

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: November 14, 2003, 23:56:24 ; Search time 949.922 Seconds  
(without alignments)  
9241.609 Million cell updates/sec

Title: US-09-928-457-67  
Perfect score: 192  
Sequence: 1 GATCCGCGTACTTGCTTTT.....CCGTTACCGACCTTCGAGA 192

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2889711 seqs, 20454813386 residues  
Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Result No.	Score	Query Match %	Length	DB ID	Description
1	192	100.0	192	6	A68896 Sequence 67
2	192	100.0	192	6	BD062996 DNA and s
3	162.4	84.6	229	1	AF169453 Neisseria
4	162.4	84.6	229	6	A68908 Sequence 79
5	162.4	84.6	229	6	BD063008 DNA and s
6	162.4	84.6	15053	1	AE002550 Neisseria
7	162.4	84.6	172325	6	AX044035 Sequence
8	162.4	84.6	349061	1	NMA222491
9	162.4	84.6	349980	6	AX044034 Sequence
10	160.8	83.8	349	1	AF169440
11	152.8	79.6	400	1	AF169418
12	118.8	61.9	409	1	NMA35415
13	65.4	34.1	1275	6	A96274 Sequence 30
14	65.4	34.1	1572	6	AX024062
15	65.4	34.1	1575	6	A96272
16	65.4	34.1	1576	6	A96270
17	60.6	31.6	1575	6	A96276 Sequence 30
18	60.6	31.6	1575	6	AX024113 Sequence
19	36.2	18.9	152572	2	AC090588
20	36.2	18.9	152881	2	AC085737
21	36.2	18.9	183169	9	AC009643
22	35.4	18.4	159539	9	AL512625 Human DNA
23	35.4	18.4	173690	2	AC024055
24	34.6	18.0	196006	10	AL645973 Mouse DNA
25	34.4	17.9	123004	9	AC087892 Homo sapi
26	34.4	17.9	175733	2	AC081621
27	34.4	17.9	177162	2	AC086796
28	34.4	17.9	188361	5	AC087880
29	34.2	17.8	137705	5	AL929326 Zebrafish
30	34.2	17.8	180594	10	AC112270
31	34	17.7	202267	9	DJ293M10
32	34	17.7	220895	10	AC125081
33	33.6	17.5	855239	9	AC026339
34	33.6	17.5	206030	9	AC021439 Homo sapi
35	33.4	17.4	232297	2	AC097198 Rattus no
36	32.8	17.1	1603	3	FLPAP12
37	32.8	17.1	74586	6	AX356520
38	32.8	17.1	74586	6	AX522013 Sequence
39	32.8	17.1	78160	9	AL136102 Human DNA
40	32.8	17.1	143145	9	AL591377 Human DNA
41	32.8	17.1	159817	2	BX322798 Danio rer
42	32.8	17.1	161218	5	AL772148 Zebrafish
43	32.8	17.1	173854	2	AL137023 Homo sapi
44	32.8	17.1	176361	2	BX322785 Danio rer
45	32.8	17.1	192846	2	AC020585 Homo sapi

ALIGNMENTS

RESULT 1  
A68896 A68896 Sequence 67 from Patent WO9802547; 192 bp DNA linear PAT 06-MAY-1999  
LOCUS  
DEFINITION  
ACCESSION A68896  
VERSION A68896.1 GI:4759815  
KEYWORDS  
SOURCE unidentified  
ORGANISM unclassified.  
REFERENCE 1 (bases 1 to 192)  
AUTHORS Nassif,X., Tinsley,C., Achtman,M., Ruelle,J., Vinals,C. and  
Marker,P.  
TITLE  
DNA AND SPECIFIC PROTEINS OR PEPTIDES OF THE NEISSERIA MENINGITIDIS  
SPECIES BACTERIA, METHOD FOR OBTAINING THEM AND THEIR BIOLOGICAL

Pred. No. is the number of results predicted by chance to have a

APPLICATIONS  
 JOURNAL Patent: WO 9802547-A 67 22-JAN-1998;  
 INST NAT SANTE RECH MED (FR)  
 COMMENT Other publication FR 2751000 19980116.  
 FEATURES Location/Qualifiers  
 source  
 1..192  
 /organism="unidentified"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:32644"  
 BASE COUNT 38 a 51 c 39 g 64 t  
 ORIGIN

Query Match 100.0%; Score 192; DB 6; Length 192;  
 Best Local Similarity 100.0%; Pred. No. 2e-49;  
 Matches 192; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATCGCGTACTGGTTTTCATATTTGCATAGTCTTGTGGTGGGCACTCTCCCGA 60  
 |||||  
 Db 1 GATCGCGTACTGGTTTTCATATTTGCATAGTCTTGTGGTGGGCACTCTCCCGA 60  
 |||||

QY 61 CATCATCTAAATTTGCTTTTATGTTTACGCCACTCATTCGGGATAAACAATATTC 120  
 |||||  
 Db 61 CATCATCTAAATTTGCTTTTATGTTTACGCCACTCATTCGGGATAAACAATATTC 120  
 |||||

QY 121 GCCTTGGCGTCGGAATGTTCAAGCTAGCTGCATCAGCTTAATCAGGTTGCCGTTACC 180  
 |||||  
 Db 121 GCCTTGGCGTCGGAATGTTCAAGCTAGCTGCATCAGCTTAATCAGGTTGCCGTTACC 180  
 |||||

QY 181 GAGCCTTCGAGA 192  
 |||||  
 Db 181 GAGCCTTCGAGA 192  
 |||||

RESULT 2  
 BD062996 122 bp DNA linear PAT 27-AUG-2002  
 LOCUS  
 DEFINITION DNA and specific proteins or peptides of the *Neisseria meningitidis* species bacteria, method for obtaining them and their biological applications.  
 ACCESSION BD062996  
 VERSION BD062996.1 GI:22608599  
 KEYWORDS JP 2001504684-A/58.  
 SOURCE unidentified  
 ORGANISM unidentified  
 unclassified.

REFERENCE 1 (bases 1 to 192)  
 AUTHORS Nassif, X., Tinsley, C., Achtman, M., Ruelle, J.L., Vinals, C. and Merker, P.  
 TITLE DNA and specific proteins or peptides of the *Neisseria meningitidis* species bacteria, method for obtaining them and their biological applications

JOURNAL  
 Patent: JP 2001504684-A 58 10-APR-2001;  
 INSTITUT NATIONAL DE LA SANTE ET DE LA RECHERCHE MEDICALE, MAX PLANCK GESELLSCHAFT ZUR FORDERUNG DER WISSENSCHAFTEN EV BERLIN, SMITHKLINE BEECHAM  
 PN JP 2001504684-A/58  
 PD 10-APR-2001  
 PF 11-JUL-1997 JP 19980505685  
 PR 12-JUL-1996 FR 96/08768  
 PI XAVIER NASSIF, COLIN TINSLEY, MARK ACHTMAN, JEAN LOUIS RUELLE, PI CARLA VINALS,  
 PI PETRA MERKER  
 PC C12N15/31, C07K14/22, C07K16/12, A61K39/095, C12Q1/68, G01N33/53 CC  
 Strandedness: Single;  
 CC Topology: Linear;  
 PH Key Location/Qualifiers.

FEATURES  
 source  
 1..192  
 /organism="unidentified"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:32644"  
 BASE COUNT 38 a 51 c 39 g 64 t  
 ORIGIN

Query Match 100.0%; Score 192; DB 6; Length 192;  
 Best Local Similarity 100.0%; Pred. No. 2e-49;  
 Matches 192; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATCGCGTACTGGTTTTCATATTTGCATAGTCTTGTGGTGGGCACTCTCCCGA 60  
 |||||  
 Db 1 GATCGCGTACTGGTTTTCATATTTGCATAGTCTTGTGGTGGGCACTCTCCCGA 60  
 |||||

QY 61 CATCATCTAAATTTGCTTTTATGTTTACGCCACTCATTCGGGATAAACAATATTC 120  
 |||||  
 Db 61 CATCATCTAAATTTGCTTTTATGTTTACGCCACTCATTCGGGATAAACAATATTC 120  
 |||||

QY 121 GCCTTGGCGTCGGAATGTTCAAGCTAGCTGCATCAGCTTAATCAGGTTGCCGTTACC 180  
 |||||  
 Db 121 GCCTTGGCGTCGGAATGTTCAAGCTAGCTGCATCAGCTTAATCAGGTTGCCGTTACC 180  
 |||||

QY 181 GAGCCTTCGAGA 192  
 |||||  
 Db 181 GAGCCTTCGAGA 192  
 |||||

RESULT 3  
 AF169453/c  
 LOCUS  
 DEFINITION *Neisseria meningitidis* strain Z249; clone Cm043 unknown sequence.  
 ACCESSION AF169453  
 VERSION AF169453.1 GI:9754661  
 KEYWORDS  
 SOURCE  
 ORGANISM *Neisseria meningitidis*  
*Neisseria meningitidis*  
 Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
 Neisseriaceae; *Neisseria*.  
 1 (bases 1 to 229)  
 PERRIN, A., NASSIF, X. and TINSLEY, C.R.  
 IDENTIFICATION OF REGIONS OF THE CHROMOSOME OF *NEISSERIA MENINGITIDIS* AND *NEISSERIA GONORRHOEA* WHICH ARE SPECIFIC TO PATHOGENIC *NEISSERIAE*  
 UNPUBLISHED  
 REFERENCE 2 (bases 1 to 229)  
 AUTHORS PERRIN, A., NASSIF, X. and TINSLEY, C.R.  
 TITLE Direct Submission  
 JOURNAL Submitted (16-JUL-1999) Necker Medicine Faculty, INSERM U411, 156 rue de Vaugirard, Paris 75015, France  
 FEATURES  
 Location/Qualifiers  
 source  
 1..229  
 /organism="Neisseria meningitidis"  
 /mol\_type="genomic DNA"  
 /strain="Z2491"  
 /db\_xref="taxon:487"  
 /clone="Cm043"  
 BASE COUNT 72 a 48 c 65 g 44 t  
 ORIGIN

Query Match 84.6%; Score 162.4; DB 1; Length 229;  
 Best Local Similarity 93.8%; Pred. No. 4.3e-40;  
 Matches 180; Conservative 0; Mismatches 11; Indels 1; Gaps 1;

QY 1 GATCGCGTACTGGTTTTCATATTTGCATAGTCTTGTGGTGGGCACTCTCCCGA 60  
 |||||  
 Db 204 GATCGCGTACTGGTTTTCATATTTGCATAGTCTTGTGGTGGGCACTCTCCCGA 145  
 |||||

QY 61 CATCATCTAAATTTGCTTTTATGTTTACGCCACTCATTCGGGATAAACAATATTC 120  
 |||||  
 Db 144 CATCATCTAAATTTGCTTTTATGTTTACGCCACTCATTCGGGATAAAC-ATATTC 86  
 |||||

QY 121 GCCTTGGCGTCGGAATGTTCAAGCTAGCTGCATCAGCTTAATCAGGTTGCCGTTACC 180  
 |||||  
 Db 85 GCCTTGGCGTCGGAATGTTCAAGCTAGCTGCATCAGCTTAATCAGGTTGCCGTTACC 26  
 |||||

QY 181 GAGCCTTCGAGA 192  
 |||||  
 Db 25 GAGCCTTCGAGA 14  
 |||||



RESULT 4  
 A68908/c  
 LOCUS  
 DEFINITION  
 A68908 Sequence 79 from Patent WO9802547.  
 ACCESSION  
 A68908  
 VERSION  
 A68908.1 GI:4759827  
 KEYWORDS  
 unclassified  
 SOURCE  
 ORGANISM  
 unclassified  
 unclassified  
 REFERENCE  
 1 (bases 1 to 229)  
 AUTHORS  
 Nassif, X., Tinsley, C., Achtman, M., Ruelle, J., Vinals, C. and Merker, P.  
 TITLE  
 'DNA AND SPECIFIC PROTEINS OR PEPTIDES OF THE NEISSERIA MENINGITIDIS SPECIES BACTERIA, METHOD FOR OBTAINING THEM AND THEIR BIOLOGICAL APPLICATIONS'  
 JOURNAL  
 Patent: WO 9802547-A 79 22-JAN-1998;  
 INST NAT SANTE RECH MED (FR)  
 COMMENT  
 Other publication FR 2751000 19980116.  
 FEATURES  
 source  
 1..229  
 /organism="unidentified"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:32644"  
 BASE COUNT 72 a 48 c 65 g 44 t  
 ORIGIN  
 Query Match 84.6%; Score 162.4; DB 6; Length 229;  
 Best Local Similarity 93.8%; Pred. No. 4.3e-40;  
 Matches 180; Conservative 0; Mismatches 11; Indels 1; Gaps 1;  
 QY 1 GATCCGCGTACTGGTTTTCATATTTTCATAGTCTTGGTGGGCAATCTTCCCGA 60  
 Db 204 GATCCGCGTACTGGTTTTCATATTTTCATAGTCTTGGTGGGCAATCTTCCCGA 145  
 QY 61 CATCATCTAAATTTGCTTTTATGTTTTCATATTTTCATAGTCTTGGTGGGCAATCTTCCCGA 120  
 Db 144 CATCATCTAAATTTGCTTTTATGTTTTCATATTTTCATAGTCTTGGTGGGCAATCTTCCCGA 86  
 QY 121 GCCTTCCGTCGCGATGTTCAAGTACGCTGCATCACCGTATCAGGTTGCCCGTTACC 180  
 Db 85 GCCTTCCGTCGCGAATGTTCAAGTACGCTGCATCACCGTATCAGGTTGCCCGTTACC 26  
 QY 181 GAGCCTTCGAGA 192  
 Db 25 GAGCCTTCGCGA 14  
 RESULT 5  
 BD063008/c  
 LOCUS  
 DEFINITION  
 BD063008 DNA and specific proteins or peptides of the Neisseria meningitidis species bacteria, method for obtaining them and their biological applications.  
 ACCESSION  
 BD063008  
 VERSION  
 BD063008.1 GI:22608611  
 KEYWORDS  
 JP 2001504684-A/70.  
 SOURCE  
 unclassified  
 ORGANISM  
 unclassified  
 REFERENCE  
 1 (bases 1 to 229)  
 AUTHORS  
 Nassif, X., Tinsley, C., Achtman, M., Ruelle, J.L., Vinals, C. and Merker, P.  
 TITLE  
 DNA and specific proteins or peptides of the Neisseria meningitidis species bacteria, method for obtaining them and their biological applications  
 JOURNAL  
 Patent: JP 2001504684-A 70 10-APR-2001;  
 INSTITUT NATIONAL DE LA SANTE ET DE LA RECHERCHE MEDICALE, MAX PLANCK GESELLSCHAFT ZUR FORDERUNG DER WISSENSCHAFTEN EV BERLIN, SMITHKLINE BEECHAM  
 COMMENT  
 PN JP 2001504684-A/70

PD 10-APR-2001  
 PF 11-JUL-1997 JP 19980505685  
 PR 12-JUL-1996 FR 96/08768  
 PI XAVIER NASSIF, COLIN TINSLEY, MARK ACHTMAN, JEAN LOUIS RUELLE, PI CARIA VINALS,  
 PI PETRA MERKER  
 PC C12N15/31, C07K14/22, C07K16/12, A61K39/095, C12Q1/68, G01N33/53 CC  
 Strandedness: Single;  
 CC Topology: Linear;  
 FH Key Location/Qualifiers.  
 1..229  
 /organism="unidentified"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:32644"  
 BASE COUNT 72 a 48 c 65 g 44 t  
 ORIGIN  
 Query Match 84.6%; Score 162.4; DB 6; Length 229;  
 Best Local Similarity 93.8%; Pred. No. 4.3e-40;  
 Matches 180; Conservative 0; Mismatches 11; Indels 1; Gaps 1;  
 QY 1 GATCCGCGTACTGGTTTTCATATTTTCATAGTCTTGGTGGGCAATCTTCCCGA 60  
 Db 204 GATCCGCGTACTGGTTTTCATATTTTCATAGTCTTGGTGGGCAATCTTCCCGA 145  
 QY 61 CATCATCTAAATTTGCTTTTATGTTTTCATAGTCTTGGTGGGCAATCTTCCCGA 120  
 Db 144 CATCATCTAAATTTGCTTTTATGTTTTCATAGTCTTGGTGGGCAATCTTCCCGA 86  
 QY 121 GCCTTCCGTCGCGAATGTTCAAGTACGCTGCATCACCGTATCAGGTTGCCCGTTACC 180  
 Db 85 GCCTTCCGTCGCGAATGTTCAAGTACGCTGCATCACCGTATCAGGTTGCCCGTTACC 26  
 QY 181 GAGCCTTCGAGA 192  
 Db 25 GAGCCTTCGCGA 14  
 RESULT 6  
 AE002550/c  
 LOCUS  
 DEFINITION  
 AE002550 Neisseria meningitidis serogroup B strain MC58 section 192 of 206 of the complete genome.  
 ACCESSION  
 AE002550 AE002098  
 VERSION  
 AE002550.2 GI:7413476  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Neisseria meningitidis MC58  
 Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
 Neisseriaceae; Neisseria.  
 1 (bases 1 to 15053)  
 Tettelin, H., Saunders, N.J., Heidelberg, J., Jeffries, A.C., Nelson, K.E., Eisen, J.A., Ketchum, K.A., Hood, D.W., Peden, J.F., Dodson, R.J., Nelson, W.C., Gwinn, M.L., DeBoy, R., Peterson, J.D., Hickey, E.K., Haft, D.H., Salzberg, S.L., White, O., Fleischmann, R.D., Dougherty, B.A., Mason, T., Ciecko, A., Parksey, D.S., Blair, E., Cittone, H., Clark, E.B., Cotton, M.D., Utterback, T.R., Khouri, H., Qin, H., Vamathevan, J., Gill, J., Scarlato, V., Masignani, V., Pizza, M., Grandi, G., Sun, L., Smith, H.O., Fraser, C.M., Moxon, E.R., Rappuoli, R. and Venter, J.C.  
 Complete genome sequence of Neisseria meningitidis serogroup B strain MC58  
 Science 287 (5459), 1809-1815 (2000)  
 JOURNAL  
 MEDLINE  
 20175755  
 PUBMED  
 10710307  
 REFERENCE  
 2 (bases 1 to 15053)  
 AUTHORS  
 Tettelin, H., Saunders, N.J., Heidelberg, J., Jeffries, A.C., Nelson, K.E., Eisen, J.A., Ketchum, K.A., Hood, D.W., Peden, J.F., Dodson, R.J., Nelson, W.C., Gwinn, M.L., DeBoy, R., Peterson, J.D., Hickey, E.K., Haft, D.H., Salzberg, S.L., White, O., Fleischmann, R.D., Dougherty, B.A., Mason, T., Ciecko, A., Parksey, D.S., Blair, E., Cittone, H., Clark, E.B., Cotton, M.D., Utterback, T.R., Khouri, H., Clark, E.B., Cotton, M.D., Utterback, T.R., Khouri, H.,

Qin, H., Vamathevan, J., Gill, J., Scarlato, V., Masignani, V.,  
 Pizza, M., Grandi, G., Sun, L., Smith, H.O., Fraser, C.M., Moxon, E.R.,  
 Rappuoli, R. and Venter, J.C.  
 Direct Submission  
 Submitted (17-MAR-2000) The Institute for Genomic Research, 9712  
 Medical Center Dr, Rockville, MD 20850, USA  
 COMMENT  
 On Apr 4, 2000 this sequence version replaced gi:7227263.  
 FEATURES  
 Location/Qualifiers  
 1..15053  
 /organism="Neisseria meningitidis MC58"  
 /mol\_type="genomic DNA"  
 /strain="MC58"  
 /db\_xref="taxon:122586"  
 /note="serogroup: B"  
 complement(74..982)  
 /gene="NMB2000"  
 complement(74..982)  
 /gene="NMB2000"  
 /note="conserved hypothetical protein; identified by  
 Glimmer2; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="AAF42327.1"  
 /db\_xref="GI:7227264"  
 /translation="MQTAINRADVTRFIEDMPVRLHVLNVMQHIWQKXNPA  
 AIRRALGELLAAAGVLLSGNLNEGTLIVQVQGRGLKMLVAEASDRTVATARDDET  
 AIADDESGLDGLSGGVFLTLQPKDEGEHWQGVVPLECGGIAQVLMVNMKRSQLDT  
 HIVLSASDEAAGLLVORLPEVLDEEAWHGVSTLARTLTAEELAGLDAQHLVRLPH  
 ETPRVFPEPTPEFSCTSRGKVSMDMLMLGEEVGVGVVQSGSIEVDCDFCHSKYVF  
 DETDYNALFGEDEVGVAKGLPHVTQ"  
 1227..1982  
 /gene="NMB2001"  
 1227..1982  
 /gene="NMB2001"  
 /note="conserved hypothetical protein; identified by  
 Glimmer2; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="AAF42328.1"  
 /db\_xref="GI:7227265"  
 /translation="WDSFFKPAVAVLWMPFAVRPALADELTNLLSREQILROFAD  
 EQVLPINRAPARRAGNDELIGSAMGLNEQVPLVNRVARRAGNDELIGSAMGLN  
 EQVLPINRAPARRAGNDELIGSAMGLIGIAYPGCTSVSTGDCSGFWOHLFKRAM  
 GINLPTSAEQRMGTIPVARELQDQWVFFRTLGGRISRISVGLYIGNNRFIHAPRTG  
 KNIETISJSHKYWSGKYAFARRVKNDPSRFLN"  
 2001..2219  
 /gene="NMB2002"  
 2001..2219  
 /gene="NMB2002"  
 /note="hypothetical protein; identified by Glimmer2;  
 putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="AAF42329.1"  
 /db\_xref="GI:7227266"  
 /translation="MSMEMPKNWDDQIVSCTEKVKVMSENMAELYQTAQDAFEDA  
 LLMGGERQLRDYLLALIEGLENPYRKV"  
 2364..2711  
 /gene="NMB2003"  
 2364..2711  
 /gene="NMB2003"  
 /note="conserved hypothetical protein; identified by  
 Glimmer2; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="AAF42330.1"  
 /db\_xref="GI:7227267"  
 /translation="MNIIRALLIILGCLATGETAVFLAGIKLPGSIVGMGVLPALLOA  
 GWKTSMLQQLTDALMSNLTLFLPPCVAVISYLDLTDADDWFSILVSASATLCLV

TKVHRWIRGIIR"  
 2708..3400  
 /gene="NMB2004"  
 2708..3400  
 /gene="NMB2004"  
 /note="conserved hypothetical protein; identified by  
 Glimmer2; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="AAF42331.1"  
 /db\_xref="GI:7227288"  
 /translation="MNEILRQPSVLLFLTLAVYALIVTRTGNIFCNPLVLSIVL  
 IAYLKILGIDYAVYHNAQFIDFWLKPVAVLAVPLVQNRKIFNQWLPFVIVSQAQS  
 VTGIVTGMVFAKWLGAEREWLSKSVTNPIAIEITRSIGGIPAITAATVIAAGLV  
 GOIAGYKMLKNTVMPSVSMGLTASHAMGIAASLERSRRVAAYAGLGLTENGVLTA  
 LIAPLLIPVLGF"  
 3467..4687  
 /gene="NMB2005"  
 3467..4687  
 /gene="NMB2005"  
 /note="similar to GB:M65216 SP:P38434 GB:S85363 PID:150243  
 PID:246444 percent identity: 98.77; identified by sequence  
 similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="glutamate N-acetyltransferase/ amino-acid  
 acetyltransferase"  
 /protein\_id="AAF42332.1"  
 /db\_xref="GI:7227269"  
 /translation="KAVNLTAKTAEQLPDIDGIALYTAQGVKKPKGHTDLTLLIAVAA  
 STGVAFITINRCAPVHIAKSHLDEDOGVRAIVNTGNAGTCAQGRIDALVCAA  
 AARQICKKNQVLPFSTGVLLEPLPADKIIAALPKMQPAFWNEARAIMIDIVPRAA  
 SREGKVDKHTVTRGIAKSGMHPHNMATMIGFIATDAKVSPVQLMTQIEIATF  
 NIITVDGDTSTNDSFVITATKNSQSEIDNIADPRYQOLKELLCALAEQAIVRDS  
 EGATREITVRNEDRDEARQAAVARSPVKTAFASDNLGRLLAAIGYAGVAD  
 LSHGVYSINADYRS"  
 4747..6156  
 /gene="NMB2006"  
 4747..6156  
 /gene="NMB2006"  
 /note="similar to PID:1652876 percent identity: 52.04;  
 identified by sequence similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="chloride channel protein-related protein"  
 /protein\_id="AAF42333.1"  
 /db\_xref="GI:7227270"  
 /translation="MTSTFPRRLARKIRQTRRLSRKSIAPFLLAGSALVALTALFFA  
 HADPALELNKLVOQYPWFVAWVPLGLPLIAWLTRKFAPTAGSGIPQVIAISLPL  
 YGAKTRILRLQOTLLKPLTLFLGMLFGASIGREGPSVQVGAVMGAMCKKHLA  
 FQMGENDLMAAGACGLAAAFNAPLAGVIFAIELRGIMLRWERQILLGVLSGFI  
 QVAIQNNPYFSGNGVLEHIFLWVSLGVCGAAGGLFGRLLIRGAAPAPRKIRG  
 FINEPRLIALMLGLLALLGTGYQKTYGTCYHEAQAALHGIYAPFLGAAPKWIAT  
 VFSYAGVPGGFTPLSTTIGAVLGHIAAIDISQANIILICMAAFLAGATQSPIT  
 SAVVMVPGQGLLFWMLIACIFASQVSQSPRPFPYHSGMRFRQRVLQETAQTG  
 NAPARQQTANSKTMPSFN"  
 6346..7737  
 /gene="NMB2007"  
 /note="This region contains a match to at least one other  
 gene that is not full length, and is not the result of a  
 sequencing artifact; similar to GP:1742299 percent  
 identity: 79.36; identified by sequence similarity;  
 putative"  
 7881..9524  
 /gene="NMB2008"  
 7881..9524  
 /gene="NMB2008"  
 /note="similar to PID:862629 percent identity: 53.03;  
 identified by sequence similarity; putative"  
 /codon\_start=1  
 /transl\_table=11

```
/product="ABC transporter, ATP-binding protein-related
protein".
/protein_id="AAF42334.1"
/db_xref="GI:1727271"
/translation="MDEIQPKVQLQLENIKVLKSGSTIIIVGANGTGKRLAV
YIEQLEKARHISAFRAKLNPVNVKIPKSAKTYLSYQWMDGIDVSNRKNYWDN
NSYHLNFDMLQYLAQONNIAVANNKRNKNTKIDILQEAWEITLLPH
RKLHITADIOVASVNEELYASNNMDSGERALFYLGOVLDGSGVLIPEPELHI
KHSIISLNADKIEELPDCSFLIICHIDIFATRAVAKXVIRNYPTPMDVISEVPE
NPBEETITWLGSRKELPVEGNNSLDIATRYCYPDWTIIPKGRCKDVIOVSLSLK
KSNEMPLNLCGSDVLDSDRDERIEQNLNGLIYLPVSEIENLFSITDVAKILK
LNGSDLELNKLNGKFSKELIYIDNELKDXLDFEVRKQKIDNLYLNKIDLSKI
TSDMKSLINEISTEOKIETWISIEIKNEIQRCIEQDLDKLLTIYDNKGLLAKSA
CIVLGMNRNKHPEFSWIRTLKGRNKDFIDAIRQKFLID"
9568..9750
/gene="NMB2009"
/notes="This region contains a gene with one or more
premature stops or frame shifts, and is not the result of a
sequencing artifact; similar to SP:Q09530 PID:733602
percent identity: 61.70; identified by sequence
similarity; putative"
9962..11539
gene
Query Match 84.6%; Score 162.4; DB 1; Length 15053;
Best Local Similarity 93.8%; Pred. No. 5.7e-40;
Matches 180; Conservative 0; Mismatches 11; Indels 1; Gaps 1;
QY 1 GATCCGGTACTTGGTTTTCATATTGTCATAGTCTTGCGTGGGCAATCTCCCGA 60
Db 11636 GATCCGGTACTTGGTTTTCATATTGTCATAGTCTTGCGTGGGCAATCTCCCGA 11577
QY 61 CATCATCTAAATTTGCTTTTATGTTTACGCCACTCATTCGGGATAAACAATATCC 120
Db 11576 CATCATCTAAATTTGCTTTTATGTTTACGCCACTCATTCGGGATAAACA-ATATTC 11518
QY 121 GCCTTCGGTCGGATGTTTCAGCTAGCTGATCACCCTGATCAGGTTGCCGTACC 180
Db 11517 GCCTTCGGTCGGATGTTTCAGCTGCTGATCACCCTGATCAGGTTGCCGTACC 11458
QY 181 GAGCCTTCGAGA 192
Db 11457 GAGCCTTCGAGA 11446
RESULT 7
AX044035/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Neisseria meningitidis
Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
Neisseriaceae; Neisseria.
REFERENCE
1
AUTHORS
Pizza, M., Hickey, E., Peterson, J., Tettelin, H., Venter, J.C.,
Maignan, V., Galicci, C., Mora, M., Ratti, G., Scarcelli, M.,
Scariato, V., Rappelli, R., Frazer, C.M. and Grandi, G.
Neisseria genomic sequences and methods of their use
Patent: WO 0066791-A 114 09-NOV-2000;
CHIRON CORPORATION (US); THE INSITUTE FOR GENOMIC RESEARCH (US)
FEATURES
source
1..172325
/organism="Neisseria meningitidis"
/mol_type="genomic DNA"
/db_xref="taxon:487"
/notes="sequence too long, cut in 8 pieces.-seq 1: 1 to
349980 349980 bases-seq 108: 300001 to 649980 349980
bases-seq 109: 600001 to 949980 349980 bases-seq 110:
900001 to 1249980 349980 bases-seq 111: 1200001 to 1549980
349980 bases-seq 112: 1500001 to 1849980 349980 bases-seq
113: 1800001 to 2149980 349980 bases-seq 114: 2100001 to
```

```
BASE COUNT 2272325 172325 bases"
43072 a 47583 c 41465 g 40205 t
ORIGIN
Query Match 84.6%; Score 162.4; DB 6; Length 172325;
Best Local Similarity 93.8%; Pred. No. 6.8e-40;
Matches 180; Conservative 0; Mismatches 11; Indels 1; Gaps 1;
QY 1 GATCCGGTACTTGGTTTTCATATTGTCATAGTCTTGCGTGGGCAATCTCCCGA 60
Db 24748 GATCCGGTACTTGGTTTTCATATTGTCATAGTCTTGCGTGGGCAATCTCCCGA 24689
QY 61 CATCATCTAAATTTGCTTTTATGTTTACGCCACTCATTCGGGATAAACAATATCC 120
Db 24688 CATCATCTAAATTTGCTTTTATGTTTACGCCACTCATTCGGGATAAACA-ATATTC 24630
QY 121 GCCTTCGGTCGGAAATGTTCAAGTAGCTGATCACCCTGATCAGGTTGCCGTACC 180
Db 24629 GCCTTCGGTCGGAAATGTTCAAGTAGCTGATCACCCTGATCAGGTTGCCGTACC 24570
QY 181 GAGCCTTCGAGA 192
Db 24569 GAGCCTTCGAGA 24558
RESULT 8
NMA222491
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Neisseria meningitidis Z2491
Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
Neisseriaceae; Neisseria.
REFERENCE
1
AUTHORS
Parkhill, J., Achtman, M., James, K.D., Bentley, S.D., Churcher, C.,
Klee, S.R., Morelli, G., Basham, D., Brown, D., Chillingworth, T.,
Davies, R.M., Davis, P., Devlin, K., Feltwell, T., Hamlin, N.,
Holroyd, S., Jagsels, K., Leather, S., Moule, S., Mungall, K.,
Quail, M.A., Rajandream, M.A., Rutherford, K.M., Simmonds, M.,
Skellton, J., Whitehead, S., Spratt, B.G. and Barrall, B.G.
Complete DNA sequence of a serogroup A strain of Neisseria
meningitidis Z2491
Nature 404 (6777), 502-506 (2000)
JOURNAL
MEDLINE
PUBMED
20222556
10761319
REFERENCE
2 (bases 1 to 349061)
Parkhill, J.
Direct Submission
Submitted (30-MAR-2000) Submitted on behalf of the Neisseria
sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk
Notes:
Details of N. meningitidis sequencing at the Sanger Centre are
available on the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/N\_meningitidis/).
FEATURES
source
1..349061
/organism="Neisseria meningitidis Z2491"
/mol_type="genomic DNA"
/strain="Z2491"
/db_xref="taxon:122587"
/notes="serogroup: A"
complement(24..206)
repeat_unit
310..39"
/label=ATR
209..212
220..661
RBS
gene
/gene="NMA0368"
```

```

CDS
220..681
/gene="NMA0368"
/notes="NMA0368, probable integral membrane protein, len: 133 aa; contains four probable transmembrane domains"
/codon_start=1
/transl_table=11
/product="putative integral membrane protein"
/protein_id="CAB83669.1"
/db_xref="GI:7379121"
/db_xref="SPTREMBL:Q9JWH7"
/translation="WCEQNRKSPFIIVMLLVSLVIAKSLNVAFLGNHSGMGLTV
LILGIFASLDIRYCAVANYVYMLAAIVLLALRKVPVHAAFLVALVAFSVKAVYV
DAAGNTSDIVRYCAGFYLMWYAAFAVASIGITFAGKNKKAASADGTKNDY"
671..1492
/gene="hemK"
/notes="hemK"
671..1492
/gene="NMA0369, hemK, HemK protein, len: 273 aa; similar to e.g. HEMK ECOLI P37186 HEMK protein (277 aa), fasta scores; E(): 0, 42.3% identity in 279 aa overlap. Contains P500092 N-6 Adenine-specific DNA methylases signature"
/codon_start=1
/transl_table=11
/product="HemK protein"
/protein_id="CAB83670.1"
/db_xref="GI:7379122"
/db_xref="SPTREMBL:Q9JWH6"
/translation="WTFDKWLGLSKPKNEARMLLQVSEYTRVOLLTRGGEEMPDEV
RORADLQRRNGEPVAVILGARFEYGRRTVNPVLIIPRTEHIVAEVLAIRLPEN
GRWDLGTSGAVATVALERPDAPVRASDISPPALETARKNAADLGRVFAFGSWF
DIDMSEGRKWDIVSNPIFIENGDKHLQGDURFEPFQAUDFSGLSCITLQAQAP
DRLAGGFLLEHGFDDQGAARVGLAENGFSGVETPLDPLADLRVTLGKMKHLK"
831..840
/gene="hemK"
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
1162..1171
/gene="hemK"
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
1187..1207
/gene="hemK"
/notes="P500092 N-6 Adenine-specific DNA methylases signature"
complement(1279..1288)
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
1565..1568
RBS
1578..2966
/gene="NMA0370"
1578..2966
/gene="NMA0370, probable integral membrane protein, len: 462 aa; similar to hypothetical proteins e.g. Y325_HAEIN P44640 hypothetical protein H10325 (450 aa), fasta scores; E(): 0, 49.8% identity in 464 aa overlap"
/codon_start=1
/transl_table=11
/product="putative integral membrane protein"
/protein_id="CAB83671.1"
/db_xref="GI:7379123"
/db_xref="SPTREMBL:Q9JWH5"
/translation="NNVVAVIVMLVLSRVHVVLITLITGAFVGGAVAGMPLQNTA
DAAGVSOAGIIPVFNKGLEGKALSYAMAFAMALITGSLPQOLAGAVAKLNR
GMPDSVRGSEVYKLLSILVGMQSONIPIHIAFIPWIPVPLLVNRLKIDR
RLIACVITGLVITFLPYGFGALFNEILLGNHSAAPQLDVKNINVMAMAPAL
GLAGLLAFHYRKPRLYQNAIDNADANRPQPSAYRSAAVAIAVCAIQL
MYEDSLVLGMLGFAFVGMVGNRKANDVFGEGIKVMAMVFIMIAAQFPAVMA
TGHIOPLVESMAIFGNSKMAALMVVGLLVITMGISFSFTLPFAIIVPLCVGL
GSPATVAIVGTAGALGAGSPASDSTLGTPTMGLNADQDHIRDSVIPTFIYNIP
LIIAGWIAAMVL"
2967..3191
/gene="slyX"

```

---

```

2967..3191
/gene="slyX"
/notes="NMA0371, slyX, SLX protein homolog, len: 74 aa; similar to SLX_HAEIN P44759 SLX protein homolog (73 aa), fasta scores; E(): 0.0017, 33.8% identity in 74 aa overlap, and SLX_ECOLI P10857 SLX protein (72 aa), fasta scores; E(): 0.24, 32.4% identity in 68 aa overlap"
/codon_start=1
/transl_table=11
/product="SLX protein homolog"
/protein_id="CAB83672.1"
/db_xref="GI:7379124"
/db_xref="SPTREMBL:Q9JWH4"
/translation="MDVAQEFHRITEIQLSALOEDVIAGLNAMVAELRQTLDDQQA
QLRLLYQKQMDRNPDAQEPYSLRDEIPPHY"
complement(3271..3423)
/gene="NMA0372"
complement(3271..3423)
/notes="NMA0372, unknown, questionable CDS, len: 50 aa"
/codon_start=1
/transl_table=11
/product="very hypothetical protein NMA0372"
/protein_id="CAB83673.1"
/db_xref="GI:7379125"
/db_xref="SPTREMBL:Q9JWH3"
/translation="MFGRLSPILGSDGFFLAVVEYVGVFCLOGKTAIFRSGGKRPFHRI
KGILSD"
3562..3571
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(3586..4356)
/gene="thiF"
complement(3586..4356)
/notes="thiF, probable ThiF protein, len: 256 aa; similar to e.g. THIF ECOLI P30138 THIF protein (251 aa), fasta scores; E(): 0_43.1% identity in 246 aa overlap, and MOEB_ECOLI P12282 molybdopterin biosynthesis MOEB protein. (249 aa), fasta scores; E(): 0, 43.9% identity in 244 aa overlap (note that N.m. does not have orthologs of any other molybdopterin biosynthesis proteins). Contains pfam match to entry PF00899 ThiF family. ThiF family"
/codon_start=1
/transl_table=11
/product="ThiF protein"
/protein_id="CAB83674.1"
/db_xref="GI:7379126"
/db_xref="SPTREMBL:Q9JWH2"
/translation="MTTTEHNDDAFLRYSRHILDEIGEQQKLSAAHILVVGCG
GLGAALPYLAASGIGTLTADSDTVLHNLQVAFDGDVGLKTEALADLRHIN
HTVDVTINEKLDGCGTLGVQAADIPLCCDNVATROAVNACVQAKTPLYSGAAR
FEQLAVYRFDLPDPSYACLDPGGSASGICSLGFSVFLVIGTQAAEALKILL
DAGEPHGRLAVYALEGGVQYFDPNPECPVCGAER"
3787..3796
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(3847..4257)
/gene="thiF"
/notes="Fram match to entry PF00899 ThiF family, ThiF family, score 186.80, E-value 4.1e-52"
complement(4418..4427)
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
4471..7224
/gene="ppc"
4471..7224
/gene="ppc"
/EC number="4.1.1.31"
/notes="NMA0374, ppc, phosphoenolpyruvate carboxylase, len: 917 aa; similar to many e.g. CAPP RHOP A32483 phosphoenolpyruvate carboxylase (EC 4.1.1.31) (936 aa), fasta scores; E(): 0, 43.3% identity in 928 aa overlap.

```

1 GATCCGGTACTGGTTTTCATATTTCATAGTCTTGTGGTGGGCGACATCTTCCCGA 60  
324748 GATCCGGTACTGGTTTTCATATTTCATAGTCTTGTGGTGGGCGACATCTTCCCGA 324689  
61 CATCATCTAAATTTGCTTTTATGGTTTTCAGCCACTCATTTGGGATAAACAATATTC 120

QY	61	CATCATCTAAATTTGCTTTATTTGGTTTACGCCACTCATTCGGGATAAACAATATTCC	120
Db	61	CAATCATTAATTTGCTTTATTTGGTTTACGCCACTCATTCGGGATAAAC-ATAATTCC	119
QY	121	GCCTTGGCGTCCGGAATGTTCAAGCTAGCTGTCATCACCGTAATCAAGTTGGCCGTTACC	180
Db	120	GCCTTGGCGTCCGGAATGTTCAAGCTGCTGCATCACCGGTAATCAAGTTGGCCGTTACC	179
QY	181	GAGCCTTCCGAGA	192
Db	180	GAGCCTTCCGCGA	191

RESULT 11	
AF169418	
LOCUS	400 bp DNA linear BCT 09-AUG-2000
DEFINITION	Neisseria gonorrhoeae strain FA1090 clone BG007 unknown sequence.
ACCESSION	AF169418
VERSION	AF169418.1 GI:9754626
KEYWORDS	
SOURCE	Neisseria gonorrhoeae
ORGANISM	Neisseria gonorrhoeae



KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
JOURNAL  
FEATURES  
source  
CDS

Neisseria meningitidis  
Neisseria meningitidis  
Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
Neisseriaceae; Neisseria.  
1  
Nassif, X., Tinsley, C., Aujame, L., Perrin, A., Rokbi, B.,  
Bouchardon, A. and Renaud, M.G.  
Patent: FR 2785293-A 5 05-MAY-2000;  
PASTEUR MERIEUX SERUMS VACC (FR)  
Location/Qualifiers  
1. .1572  
/organism="Neisseria meningitidis"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:487"  
1. .1572  
/notes="unnamed protein product"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAC08907.1"  
/db\_xref="GI:10184375"  
/translation="MKKSLFVFLYSSLLTASEIAYRVFGIETLPAKMAETFALTE  
MTAALYFARYKASRLIIAVFFAFSIIANNVHYAVQSWTGINVLMLEITEVGS  
GASMLDKLWLPALWGLVFCSLAKPRTKTHFSADILFALMLMI FVRSPTDKQEH  
GISPKPTYSRIKANYFSGVYGRVLPVOLFSLRIAPKQAPSKIGOGSVONIVLI  
MGESASAHLKLFVGYGRETSPBLTSLQADRYPIYKQSYAGFNTAVSLPSFENAIPH  
ANGLEISGSDTNMFLAKEQGVETTFYSAQAEENMALLNLIGKXWIDHLIOPTQIGY  
GNGDNPFDEKLGLFDPKINLQQRHFIVLHQGSHAPYSALLQPDQKVFGEIIVDKYD  
NTHTKDQMIQTVEQLQKPPGNLWFAYTSDHQYVRQDIYNOGTVPDSYLVPLVL  
YSSNKAVQQAQAQAFAPCEIAFHQQLSTFLIHTLGYDMFVSGREGSVTGNLITGDAG  
SLNIRDGKAEYVYVQ"

BASE COUNT 380 a 383 C 395 G 414 t  
ORIGIN

Query Match 34.1%; Score 65.4; DB 6; Length 1572;  
Best Local Similarity 87.4%; Pred. No. 1.8e-09;  
Matches 83; Conservative 0; Mismatches 11; Indels 1; Gaps 1;  
QY 98 TCATTGGGATAAACAAATATTCGGCTTCCGCTCGCGAATGTTCAAGCTACCTGCATCA 157  
Db 1572 TCATTGGGATAAAC-ATATTCGGCTTCCGCTCGCGAATGTTCAAGCTACCTGCATCA 157  
QY 158 CCGTAATCAGGTTGCCGTTACCGAGCCTTCGAGA 192  
Db 1513 CCGTAATCAGGTTGCCGTTACCGAGCCTTCGAGA 1479

RESULT 15  
A96272/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
source  
BASE COUNT  
ORIGIN

A96272  
Sequence 305 from Patent WO9924578.  
A96272  
A96272.1 GI:6780028  
unidentified  
unidentified  
unclassified.  
1  
Pizza, M., Scarlato, V., Rappuoli, R., Grandi, G. and Masignani, V.  
Neisserial antigens  
Patent: WO 9924578-A 305 20-MAY-1999;  
PIZZA MARIAGRAZIA (IT); SCARLATO VINCENZO (IT); RAPPUOLI RINO (IT);  
CHIRON SPA (IT); GRANDI GUIDO (IT); MASIGNANI VEGA (IT)  
Location/Qualifiers  
1. .1575  
/organism="unidentified"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32644"  
376 a 378 c 405 G 416 t

BASE COUNT 376 a 378 c 405 G 416 t  
ORIGIN

Query Match 34.1%; Score 65.4; DB 6; Length 1575;  
Best Local Similarity 87.4%; Pred. No. 1.8e-09;  
Matches 83; Conservative 0; Mismatches 11; Indels 1; Gaps 1;  
QY 98 TCATTGGGATAAACAAATATTCGGCTTCCGCTCGCGAATGTTCAAGCTACCTGCATCA 157  
Db 1572 TCATTGGGATAAAC-ATATTCGGCTTCCGCTCGCGAATGTTCAAGCTACCTGCATCA 157  
QY 158 CCGTAATCAGGTTGCCGTTACCGAGCCTTCGAGA 192  
Db 1516 CCGTAATCAGGTTGCCGTTACCGAGCCTTCGAGA 1482





	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	52
--	---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	----

DR WPI; 1998-110594/10.  
 XX Genes present in *Neisseria meningitidis* but not other *Neisseria*  
 PT species - and related host cells, RNA, anti-sense sequences,  
 PT polypeptide(s) and antibodies, useful for diagnosing *Neisseria*  
 PT meningitidis infection and in protective vaccines  
 XX  
 PS Example 4; Page 113; 150pp; French.  
 XX  
 CC AAV03575-606 represent sequences that are present in *Neisseria*  
 CC meningitidis and *N. gonorrhoeae* but not in *N. lactamica*, except for the  
 CC genes involved in biosynthesis of the capsule polysaccharide, *frpA* or *C*,  
 CC *opc*, *porA*, *rotamase*, sequence IC1106, *IGA* protease, *pilin*, *pilC*,  
 CC proteins which bind transferrin and opacity proteins. The DNA sequences  
 CC are responsible for the differences in pathogenicity between *N.*  
 CC meningitidis and *N. gonorrhoeae*, specifically they include the genes that  
 CC allow *N. meningitidis* to cross the blood-brain barrier. DNA sequences  
 CC common to *N. meningitidis* and *N. gonorrhoeae*, but absent from *N.*  
 CC *lactamica*, are responsible for colonisation and penetration of the  
 CC mucosa. The DNA sequences can be used to produce probes and primers, and  
 CC antibodies produced against the encoded proteins are used in standard  
 CC hybridisation/immunoassay processes for diagnosis of *N. meningitidis*  
 CC infection, particularly meningitis.  
 XX  
 SQ Sequence 192 BP; 38 A; 51 C; 39 G; 64 T; 0 other;  
 Query Match 100.0%; Score 192; DB 19; Length 192;  
 Best Local Similarity 100.0%; Pred. No. 5e-56;  
 Matches 192; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GATCCGCGACTGGTTTTCATATTTTCATAGTCTTGTGCGGCGCATCTTCCCGGA 60  
 DB 1 GATCCGCGACTGGTTTTCATATTTTCATAGTCTTGTGCGGCGCATCTTCCCGGA 60  
 QY 61 CATCATCTAAATTTGCTTTTATGTTTTCATGCGGCACCTATTGGGATAAACAATATTC 120  
 DB 61 CATCATCTAAATTTGCTTTTATGTTTTCATGCGGCACCTATTGGGATAAACAATATTC 120  
 QY 121 GCCTTGCCGTCGCGAATGTTCAAGCTAGCTGATCAGCCGTAATCAGTTGCCGTTACC 180  
 DB 121 GCCTTGCCGTCGCGAATGTTCAAGCTAGCTGATCAGCCGTAATCAGTTGCCGTTACC 180  
 QY 181 GAGCCTTCGAGA 192  
 DB 181 GAGCCTTCGAGA 192  
 RESULT 2  
 AAV03587/c  
 ID AAV03587 standard; DNA; 229 BP.  
 XX  
 AC AAV03587;  
 XX  
 XX  
 DT 22-OCT-1998 (first entry)  
 XX  
 DE *Neisseria meningitidis* DNA sequence C43.  
 XX  
 KW *N. gonorrhoeae*; *N. lactamica*; chromosome 22491; region 1; region 2;  
 KW region 3; pathogenicity; blood-brain barrier; diagnosis; infection;  
 KW meningitis; ss.  
 XX  
 OS *Neisseria meningitidis*.  
 XX  
 PN WO9802547-A2.  
 XX  
 PD 22-JAN-1998.  
 XX  
 PF 11-JUL-1997; 97WO-FR01295.  
 XX  
 PR 12-JUL-1996; 96FR-0008768.  
 XX  
 PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.  
 PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.

PA (SMIK) SMITHKLINE BEECHAM.  
 XX  
 PI Nassif X, Tinsley C, Achtman M, Merker P, Ruelle J;  
 PI Vinals C;  
 XX  
 DR WPI; 1998-110594/10.  
 XX  
 XX Genes present in *Neisseria meningitidis* but not other *Neisseria*  
 PT species - and related host cells, RNA, anti-sense sequences,  
 PT polypeptide(s) and antibodies, useful for diagnosing *Neisseria*  
 PT meningitidis infection and in protective vaccines  
 XX  
 PS Example 4; Page 119; 150pp; French.  
 XX  
 CC AAV03575-606 represent sequences that are present in *Neisseria*  
 CC meningitidis and *N. gonorrhoeae* but not in *N. lactamica*, except for the  
 CC genes involved in biosynthesis of the capsule polysaccharide, *frpA* or *C*,  
 CC *opc*, *porA*, *rotamase*, sequence IC1106, *IGA* protease, *pilin*, *pilC*,  
 CC proteins which bind transferrin and opacity proteins. The DNA sequences  
 CC are responsible for the differences in pathogenicity between *N.*  
 CC meningitidis and *N. gonorrhoeae*, specifically they include the genes that  
 CC allow *N. meningitidis* to cross the blood-brain barrier. DNA sequences  
 CC common to *N. meningitidis* and *N. gonorrhoeae*, but absent from *N.*  
 CC *lactamica*, are responsible for colonisation and penetration of the  
 CC mucosa. The DNA sequences can be used to produce probes and primers, and  
 CC antibodies produced against the encoded proteins are used in standard  
 CC hybridisation/immunoassay processes for diagnosis of *N. meningitidis*  
 CC infection, particularly meningitis.  
 XX  
 SQ Sequence 229 BP; 72 A; 48 C; 65 G; 44 T; 0 other;  
 Query Match 84.6%; Score 162.4; DB 19; Length 229;  
 Best Local Similarity 93.8%; Pred. No. 8.1e-46;  
 Matches 180; Conservative 0; Mismatches 11; Indels 1; Gaps 1;  
 QY 1 GATCCGCGACTGGTTTTCATATTTTCATAGTCTTGTGCGGCGCATCTTCCCGGA 60  
 DB 204 GATCCGCGACTGGTTTTCATATTTTCATAGTCTTGTGCGGCGCATCTTCCCGGA 145  
 QY 61 CATCATCTAAATTTGCTTTTATGTTTTCATGCGGCACCTATTGGGATAAACAATATTC 120  
 DB 144 CATCATCTAAATTTGCTTTTATGTTTTCATGCGGCACCTATTGGGATAAACAATATTC 96  
 QY 121 GCCTTGCCGTCGCGAATGTTCAAGCTAGCTGATCAGCCGTAATCAGTTGCCGTTACC 180  
 DB 85 GCCTTGCCGTCGCGAATGTTCAAGCTAGCTGATCAGCCGTAATCAGTTGCCGTTACC 26  
 QY 181 GAGCCTTCGAGA 192  
 DB 25 GAGCCTTCGAGA 14  
 RESULT 3  
 AAA81550/c  
 ID AAA81550 standard; DNA; 12438 BP.  
 XX  
 AC AAA81550;  
 XX  
 XX 04-DEC-2000 (first entry)  
 XX  
 DE *N. meningitidis* partial DNA sequence gnm\_97 SEQ ID NO:97.  
 XX  
 KW *Neisseria meningitidis*; *Neisseria gonorrhoeae*; genome; immunogenic;  
 KW antigen; vaccine; diagnosis; infection; antibacterial; identification;  
 KW *Meningococcus B*; MenB; ds.  
 XX  
 OS *Neisseria meningitidis*.  
 XX  
 PN WO200022430-A2.  
 XX  
 PD 20-APR-2000.  
 XX  
 PF 08-OCT-1999; 99WO-US23573.

DT	13-MAR-2001	(first entry)	
XX	Neisseria meningitidis B	nucleotide sequence SEQ ID NO:114.	
XX	Neisseria meningitidis	B nucleotide sequence SEQ ID NO:114.	
XX	Neisseria meningitidis	Neisseria gonorrhoeae; immunogenic; vaccine;	
KW	diagnosis; antigen; detection; infection; gene therapy; antibacterial;	ds.	
XX	Neisseria meningitidis.		
XX	WO200066791-A1.		
PN	09-NOV-2000.		
PD			
XX	08-MAR-2000; 2000WO-US05928.		
PF			
XX	30-APR-1999; 99US-0132068.		
PR	08-OCT-1999; 99WO-US23573.		
PR	28-FEB-2000; 2000GB-0004695.		
XX	(CHIR ) CHIRON CORP.		
PA	(GENO-) INST GENOMIC RES.		
PI	Pizza M, Hickey E, Peterson J, Tettelin H, Venter JC, Massignani V;		
PI	Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V, Rappuoli R;		
PI	Frazer CM, Grandi G;		
XX	WPI; 2000-647603/62.		
DR			
XX	Neisseria meningitidis B	full length genome sequence and open reading	
PT	frames are used to detect, treat and prevent Neisserial infections -		
PT	Claim 7; Appendix A; 692pp; English.		
XX	The present invention describes the full length genome of		
CC	Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607		
CC	to AAF21613 represent fragments of the NMB genomic sequence, as the		
CC	sequence was too long to go in a record on its own it was split into 8		
CC	sequences which overlap each other at the beginning and end of each		
CC	sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at		
CC	the beginning of AAF21607, the last 49980 bp of AAF21607 are repeated at		
CC	the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the		
CC	Neisseria proteins given in AAB58550 to AAB58593, and AAF21589 to		
CC	AAF21606 represent PCR primers which are used in the exemplification of		
CC	the present invention. The NMB genome and fragments from it have		
CC	antibacterial activity, and can be used in vaccines and gene therapy.		
CC	Neisseria nucleic acids, proteins and/or antibodies which binds to the		
CC	proteins can be used in compositions for treating or preventing infection		
CC	due to Neisserial bacteria or as a diagnostic reagent for detecting the		
CC	presence of Neisserial bacteria or of antibodies raised to Neisserial		
CC	bacteria. Computers, computer memory, computer storage medium or computer		
CC	databases can be used in a search to identify open reading frames (ORFs)		
CC	or coding sequences within the NMB genome. The DNA sequences provide		
CC	further opportunities to find antigenic or immunogenic proteins which are		
CC	more effective in vaccines than the outer membrane proteins currently		
XX	used.		
XX	Sequence 172325 BP; 43072 A; 47583 C; 41465 G; 40205 T; 0 other;		
SQ	Query Match	84.6%; Score 162.4; DB 21; Length 172325;	
	Best Local Similarity	93.8%; Pred. No.1.e-44;	
	Matches 180; Conservative 0; Mismatches 11; Indels 1; Gaps 1;		
QY	1	GATCGCGCTACTTGTTTTCATATTTTGCATAGTCTTGTCGGTCGGGCATCTTCCCGGA	60
Db	24748	GATCGCGCTACTTGTTTTCATATTTTGCATAGTCTTGTCGGTCGGGCATCTTCCCGGA	24589
QY	61	CATCATCTAAATTGTCCTTTATGGTTTTCAGCGCACTCATTCGGGTAACATATTC	120
Db	24688	CATCATCTAAATTGTCCTTTATGGTTTTCAGCGCACTCATTCGGGTAAC-ATATTC	24630
QY	121	GCCTTGGCGTCCGGAATGTCCTCAAGTACGCTGCATCCGTAATCAGGTTCCCGGTACC	180

Db 24629 GCCTTCCGCTCGGAATGTTCAAGCTGCTGCATCACCGGTAATCAGGTGCCGTTACC 24570

Qy 181 GAGCCTTCGAGA 192

Db 24569 GAGCCTTCGAGA 24558

# RESULT 5

ID AAF21612/c

XX AAF21612 standard; DNA; 349980 BP.

AC AAF21612;

XX 13-MAR-2001 (first entry)

XX Neisseria meningitidis B nucleotide sequence SEQ ID NO:113.

XX Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;  
KW diagnosis; antigen; detection; infection; gene therapy; antibacterial;  
KW ds.

XX Neisseria meningitidis.

OS WO200066791-A1.

PN 09-NOV-2000.

XX 08-MAR-2000; 2000WO-US05928.

XX 30-APR-1999; 99US-0132068.

PR 08-OCT-1999; 99WO-US23573.

PR 28-FEB-2000; 2000GB-0004695.

XX (CHIR ) CHIRON CORP.

PA (GENO-) INST GENOMIC RES.

XX Pizza M, Hickey E, Peterson J, Tettelin H, Venter JC, Masignani V;  
PI Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V, Rappuoli R;  
PI Frazer CM, Grandi G;

XX WPI; 2000-647603/62.

XX Neisseria meningitidis B full length genome sequence and open reading  
frames are used to detect, treat and prevent Neisserial infections -

PS Claim 7; Appendix A; 692pp; English.

XX The present invention describes the full length genome of  
CC Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607  
CC to AAF21613 represent fragments of the NMB genomic sequence, as the  
CC sequence was too long to go in a record on its own it was split into 8  
CC sequences which overlap each other at the beginning and end of each  
CC sequence by 4980 bp (i.e. the last 4980 bp of AAF21544 is repeated at  
CC the beginning of AAF21607, the last 4980 bp of AAF21607 are repeated at  
CC the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the  
CC Neisseria proteins given in AAB58550 to AAB58593, and AAF21589 to  
CC AAF21606 represent PCR primers which are used in the exemplification of  
CC the present invention. The NMB genome and fragments from it have  
CC antibacterial activity, and can be used in vaccines and gene therapy.  
CC Neisseria nucleic acids, proteins and/or antibodies which binds to the  
CC proteins can be used in compositions for treating or preventing infection  
CC due to Neisserial bacteria or as a diagnostic reagent for detecting the  
CC presence of Neisserial bacteria or of antibodies raised to Neisserial  
CC bacteria. Computers, computer memory, computer storage medium or computer  
CC databases can be used in a search to identify open reading frames (ORFs)  
CC or coding sequences within the NMB genome. The DNA sequences provide  
CC further opportunities to find antigenic or immunogenic proteins which are  
CC more effective in vaccines than the outer membrane proteins currently  
CC used.

XX Sequence 349980 BP; 86473 A; 95646 C; 85908 G; 81953 T; 0 other;

SQ Query Match 84.6%; Score 162.4; DB 21; Length 349980;

Best Local Similarity 93.8%; Pred. No. 1.5e-44;  
Matches 180; Conservative 0; Mismatches 11; Indels 1; Gaps 1;

Qy 1 GATCCGCGTACTGGTTTTCATATTTTCATAGTCTTCGTCGGGCGATCTTCCCGA 60

Db 324748 GATCCGCGTACTGGTTTTCATATTTTCATAGTCTTCGTCGGGCGATCTTCCCGA 324889

Qy 61 CATCATCTAAATTTGCTTTTATTTGTTTTACGCCACTCATTCGGGATAAACAATATTC 120

Db 324688 CATCATCTAAATTTGCTTTTATTTGTTTTACGCCACTCATTCGGGATAAACAATATTC 324630

Qy 121 GCCTTCCGCTCGGAATGTTCAAGCTAGCTGCATCACCGTATCAGGTGCCCGTTACC 180

Db 324629 GCCTTCCGCTCGGAATGTTCAAGCTAGCTGCATCACCGTATCAGGTGCCCGTTACC 324570

Qy 181 GAGCCTTCGAGA 192

Db 324569 GAGCCTTCGAGA 324558

## RESULT 6

AAA81489/c

ID AAA81489 standard; DNA; 837096 BP.

XX AAA81489;

AC AAA81489;

XX 04-DEC-2000 (first entry)

DE N. meningitidis partial DNA sequence gnm\_37 SEQ ID NO:37.

XX Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;  
KW antigen; vaccine; diagnosis; infection; antibacterial; identification;  
KW Meningococcus B; MenB; ds.

XX Neisseria meningitidis.

XX WO200022430-A2.

XX 20-APR-2000.

XX 08-OCT-1999; 99WO-US23573.

XX 09-OCT-1998; 98US-0103794.

XX 30-APR-1999; 99US-0132068.

XX (CHIR ) CHIRON CORP.

PA Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC, Scarlato V;  
PI Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;  
PI Rappuoli R, Pizza M;

XX WPI; 2000-318079/27.

XX Isolated nucleotide sequences of Neisseria meningitidis which can be  
used in the diagnosis and treatment of N. meningitidis infection and  
other Neisserial infections, for example, N.gonorrhoea -  
Claim 7; Page 629-865; 1760pp; English.

XX The present invention describes methods of obtaining immunogenic  
CC proteins from Neisseria genomic sequences. AAA81453 to AAA82414  
CC represent specifically claimed Neisseria meningitidis genomic DNA  
CC sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent  
CC Neisseria DNA sequences and their corresponding proteins; AAA81254 to  
CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the  
CC isolation of Neisseria meningitidis DNA sequences; and AAA81322 to  
CC AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF  
CC sequences, which are all used in the exemplification of the present  
CC invention. The nucleic acid sequences, protein sequences, and antibodies  
CC against them, can be used in the manufacture of a composition. The  
CC composition can be used as a medicament (or in the manufacture of a  
CC medicament) for treating, preventing or diagnosing infection due to  
CC Neisserial bacteria. For example, some of the identified proteins could

XX The present invention describes methods of obtaining immunogenic  
CC proteins from Neisseria genomic sequences. AA81453 to AAA92414  
CC represent specifically claimed Neisseria meningitidis genomic DNA  
CC sequences; AA81260 to AA81303 and AAB25620 to AAB25663 represent  
CC Neisseria DNA sequences and their corresponding proteins; AA81254 to  
CC AA81259 and AA81304 to AA81331 represent PCR primers used in the  
CC isolation of Neisseria meningitidis DNA sequences; and AA81332 to  
CC AA81452 represent Neisseria meningitidis MenB polynucleotide ORF  
CC sequences, which are all used in the exemplification of the present  
CC invention. The nucleic acid sequences, protein sequences, and antibodies  
CC against them, can be used in the manufacture of a composition. The  
CC composition can be used as a medicament (or in the manufacture of a  
CC medicament) for treating, preventing or diagnosing infection due to  
CC Neisserial bacteria. For example, some of the identified proteins could  
CC be components of vaccines against Meningococcus B; against all serotypes;  
CC and/or against all pathogenic Neisseriae. Identification of sequences  
CC from the bacterium will also facilitate production of biological probes,  
CC particularly organism-specific probes. Attempts to make efficacious  
CC Meningococcus B vaccines have failed mainly due to antigen tolerance.  
CC Multivalent vaccines have also been tried but none have successfully  
CC overcome antigenic variability. The provision of further, complete  
CC sequences may provide an opportunity to identify secreted or surface  
CC exposed proteins that may be presumed targets for the immune system and  
CC which are not antigenically variable or at least more conserved than  
CC other more variable regions.

XX  
XX SQ Sequence 366 BP; 84 A; 100 C; 97 G; 85 T; 0 other;

Query Match 34.1%; Score 65.4; DB 21; Length 366;  
Best Local Similarity 87.4%; Pred. No. 2.2e-12;  
Matches 83; Conservative 0; Mismatches 11; Indels 1; Gaps 1;

QY 98 TCATTGGGATAAACAAATATTCGCCTTCCCGTCGGGAATGTTCAGCTAGCCTGCATCA 157  
Db 366 TCATTGGGATAAAC-ATATTCGCCTTCCCGTCGGGAATGTTCAGCTAGCCTGCATCAC 308

QY 158 CCCTAATCAGTTGCCCGTTACCAGCCTTCGAGA 192  
Db 307 CCGTATCAGTTGCCCGTTACCAGCCTTCGCGA 273

RESULT 8  
AAZ12104/C  
ID ID AAZ12104 standard; DNA; 624 BP.  
XX AC AAZ12104;  
XX AC AAZ12104;  
XX DT 08-OCT-1999 (first entry)  
XX DE Neisseria meningitidis partial ORF81 sequence.  
XX KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;  
KW treatment; Neisseria infection; meningitis; septicemia, gonorrhea; ss.  
XX CS Neisseria meningitidis.  
XX SS Neisseria meningitidis.  
XX FH Key Location/Qualifiers  
FH misc\_feature 256..257  
FT FT /\*tag= a  
FT FT /note= "52 unspecified nucleotides are present  
FT between these bases"  
XX .  
XX . WO9924578-A2.  
XX XX  
XX PD 20-MAY-1999.  
XX XX  
XX PF 03-OCT-1998; 98WO-IB01565.  
XX XX  
XX PR 01-SEP-1998; 98GB-0019016.  
PR 06-NOV-1997; 97GB-0023516.  
PR 14-NOV-1997; 97GB-0024190.  
PR 18-NOV-1997; 97GB-0024386.  
PR .

```

PR 27-NOV-1997; 97GB-0025158.
PR 10-DEC-1997; 97GB-0026147.
PR 14-JAN-1998; 98GB-0000759.
XX (CHIR-) CHIRON SPA.
XX Grandi G, Masignani V, Pizza M, Rappuoli R, Scarlato V;
XX WPI; 1999-327407/27.
DR P-PSDB; AAY38650.
XX Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT diagnosis, treatment and prevention of infection
XX Claim 9; Page 204; 524pp; English.
XX Nucleotide sequences AA211972-212358 represent open reading frames
CC (ORFs) of Neisseria meningitidis and N. gonorrhoeae which encode
CC antigenic proteins (see AAY38499-Y38944). The antigenic proteins, their
CC fragments, their nucleic acids and antibodies are used for diagnosis,
CC prevention (as vaccines) or treatment of Neisseria infections,
CC such as meningitis, septicemia and gonorrhea. Both organisms
CC are closely related. Fragments of the nucleic acids are useful
CC as hybridisation probes and antisense reagents.
XX SQ Sequence 624 BP; 141 A; 152 C; 159 G; 170 T; 2 other;
Query Match 34.1%; Score 65.4; DB 20; Length 624;
Best Local Similarity 87.4%; Pred. No. 2.7e-12;
Matches 83; Conservative 0; Mismatches 11; Indels 1; Gaps 1;

QY 98 TCATTCCGGATAAACAATATTCGCGCTTCGCGTAATGTTCAAGTAGCCTGCATCA 157
Db 624 TCATTCCGGATAAAC-ATATTCGCGCTTCGCGTAATGTTCAAGTGCCTGCATCAC 566

QY 158 CCGTAATCAGGTGCGCGTTACCGAGCCTTCGAGA 192
Db 565 CCGTAATCAGGTGCGCGTTACCGAGCCTTCGCGA 531

RESULT 9
AA212106/c
ID AA212106 standard; DNA; 1275 BP.
XX AA212106;
XX 08-OCT-1999 (first entry)
XX Neisseria meningitidis strain A complete ORF81 sequence.
XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
XX treatment; Neisseria infection; meningitis; septicemia; gonorrhea; ss.
XX Neisseria meningitidis.
XX WO924578-A2.
XX 20-MAY-1999.
XX 09-OCT-1998; 98WO-IB01665.
XX 01-SEP-1998; 98GB-0019016.
XX 06-NOV-1997; 97GB-0023516.
XX 14-NOV-1997; 97GB-0024190.
XX 18-NOV-1997; 97GB-0024386.
XX 27-NOV-1997; 97GB-0025158.
XX 10-DEC-1997; 97GB-0026147.
XX 14-JAN-1998; 98GB-0000759.
XX (CHIR-) CHIRON SPA.
XX Grandi G, Masignani V, Pizza M, Rappuoli R, Scarlato V;
XX

```

```

DR WPI; 1999-327407/27.
DR P-PSDB; AAY38652.
XX Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT diagnosis, treatment and prevention of infection
XX Claim 9; Page 206; 524pp; English.
XX Nucleotide sequences AA211972-212358 represent open reading frames
CC (ORFs) of Neisseria meningitidis and N. gonorrhoeae which encode
CC antigenic proteins (see AAY38499-Y38944). The antigenic proteins, their
CC fragments, their nucleic acids and antibodies are used for diagnosis,
CC prevention (as vaccines) or treatment of Neisseria infections,
CC such as meningitis, septicemia and gonorrhea. Both organisms
CC are closely related. Fragments of the nucleic acids are useful
CC as hybridisation probes and antisense reagents.
XX SQ Sequence 1275 BP; 300 A; 304 C; 319 G; 352 T; 0 other;
Query Match 34.1%; Score 65.4; DB 20; Length 1275;
Best Local Similarity 87.4%; Pred. No. 3.6e-12;
Matches 83; Conservative 0; Mismatches 11; Indels 1; Gaps 1;

QY 98 TCATTCCGGATAAACAATATTCGCGCTTCGCGTAATGTTCAAGTAGCCTGCATCA 157
Db 1275 TCATTCCGGATAAAC-ATATTCGCGCTTCGCGTAATGTTCAAGTGCCTGCATCAC 1217

QY 158 CCGTAATCAGGTGCGCGTTACCGAGCCTTCGAGA 192
Db 1216 CCGTAATCAGGTGCGCGTTACCGAGCCTTCGCGA 1182

RESULT 10
AAAL5298/c
ID AAAL5298 standard; DNA; 1572 BP.
XX AAAL5298;
XX 04-SEP-2000 (first entry)
XX DNA encoding a polypeptide of a Neisseria pathogenic strain.
XX Pathogenic strain; Neisseria; vaccine; Neisseria infection; ss.
XX Neisseria meningitidis.
XX Key Location/Qualifiers
XX CDS 1..1572
XX FT /*tag= a
XX WO200026375-A2.
XX 11-MAY-2000.
XX 28-OCT-1999; 99WO-FR02643.
XX 30-OCT-1998; 98FR-0013693.
XX (INMR ) PASTEUR MERIEUX SERUMS & VACCINS SA.
XX (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX Aujame L, Bouchardon A, Renaud-Mongenie G, Rokbi B, Nassif X;
XX Tinsley C, Perrin A;
XX WPI; 2000-355622/31.
XX P-PSDB; AAY93267.
XX New polypeptide specific for pathogenic Neisseria useful in therapeutic
XX or preventative vaccines and for diagnosis
XX Claim 2; Page 56-59; 187pp; French.
XX The present sequence encodes a protein that is specific for pathogenic
XX

```

CC strains of Neisseria. The polynucleotides, polypeptides, or their  
 CC antigenic fragments, are used in vaccines to treat or protect against  
 CC Neisseria infections, particularly by N. meningitidis. The  
 CC polynucleotide sequence is also used for recombinant production of  
 CC the polypeptide and to produce attenuated Neisseria strains that  
 CC overexpress it, or express it in a non-toxic mutant form.  
 XX  
 SQ Sequence 1572 BP; 380 A; 383 C; 395 G; 414 T; 0 other;

Query Match 34.1%; Score 65.4; DB 21; Length 1572;  
 Best Local Similarity 87.4%; Pred. No. 3.9e-12;  
 Matches 83; Conservative 0; Mismatches 11; Indels 1; Gaps 1;  
 QY 98 TCATTGGGATAAACAAATATTCGCGCTTCGCGAATGTTCAAGCTAGCTGCATCA 157  
 Db 1572 TCATTGGGATAAAC-ATATTCGCGCTTCGCGAATGTTCAAGCTAGCTGCATCAC 1514  
 QY 158 CCGTAATCAGGTGCCCGTTACCGAGCCTTCGAGA 192  
 Db 1513 CCGTAATCAGGTGCCCGTTACCGAGCCTTCGCGA 1479

RESULT 11  
 ID AAZ12105/C  
 XX AAZ12105 standard; DNA; 1575 BP.  
 AC AAZ12105;  
 DX 08-OCT-1999 (first entry)  
 XX Neisseria meningitidis complete ORF81 sequence.  
 DE Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;  
 KW treatment; Neisseria infection; meningitis; septicaemia; gonorrhea; ss.  
 XX Neisseria meningitidis.  
 XX WO9924578-A2.  
 XX 20-MAY-1999.  
 XX 09-OCT-1998; 98WO-IB01665.  
 XX 01-SEP-1998; 98GB-0019016.  
 XX 06-NOV-1997; 97GB-0023516.  
 XX 14-NOV-1997; 97GB-0024190.  
 XX 18-NOV-1997; 97GB-0024386.  
 XX 27-NOV-1997; 97GB-0025158.  
 XX 10-DEC-1997; 97GB-0026147.  
 XX 14-JAN-1998; 98GB-0000759.  
 XX (CHIR-) CHIRON SPA.  
 XX Grandi G, Massignani V, Pizza M, Rappuoli R, Scarlato V;  
 WPI; 1999-327407/27.  
 DR P-PSDB; AAY38651.  
 XX Proteins from Neisseria meningitidis and N. gonorrhoeae useful for  
 PT diagnosis, treatment and prevention of infection  
 XX  
 PS Claim 9; Page 205; 524pp; English.  
 XX Nucleotide sequences AAZ11972-721358 represent open reading frames  
 CC (ORFs) of Neisseria meningitidis and N. gonorrhoeae which encode  
 CC antigenic proteins (see AAY38499-Y38944). The antigenic proteins, their  
 CC fragments, their nucleic acids and antibodies are used for diagnosis,  
 CC prevention (as vaccines) or treatment of Neisseria infections,  
 CC such as meningitis, septicaemia and gonorrhea. Both organisms  
 CC are closely related. Fragments of the nucleic acids are useful  
 CC as hybridisation probes and antisense reagents.  
 XX  
 SQ Sequence 1575 BP; 376 A; 378 C; 405 G; 416 T; 0 other;

Query Match 34.1%; Score 65.4; DB 20; Length 1575;  
 Best Local Similarity 87.4%; Pred. No. 3.9e-12;  
 Matches 83; Conservative 0; Mismatches 11; Indels 1; Gaps 1;  
 QY 98 TCATTGGGATAAACAAATATTCGCGCTTCGCGAATGTTCAAGCTAGCTGCATCA 157  
 Db 1575 TCATTGGGATAAAC-ATATTCGCGCTTCGCGAATGTTCAAGCTAGCTGCATCAC 1517  
 QY 158 CCGTAATCAGGTGCCCGTTACCGAGCCTTCGAGA 192  
 Db 1516 CCGTAATCAGGTGCCCGTTACCGAGCCTTCGCGA 1482

RESULT 12  
 ID AAZ12107/C  
 XX AAZ12107 standard; DNA; 1575 BP.  
 AC AAZ12107;  
 DX 08-OCT-1999 (first entry)  
 XX Neisseria gonorrhoeae complete ORF81 sequence.  
 DE Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;  
 KW treatment; Neisseria infection; meningitis; septicaemia; gonorrhea; ss.  
 XX Neisseria gonorrhoeae.  
 XX WO9924578-A2.  
 XX 20-MAY-1999.  
 XX 09-OCT-1998; 98WO-IB01665.  
 XX 01-SEP-1998; 98GB-0019016.  
 XX 06-NOV-1997; 97GB-0023516.  
 XX 14-NOV-1997; 97GB-0024190.  
 XX 18-NOV-1997; 97GB-0024386.  
 XX 27-NOV-1997; 97GB-0025158.  
 XX 10-DEC-1997; 97GB-0026147.  
 XX 14-JAN-1998; 98GB-0000759.  
 XX (CHIR-) CHIRON SPA.  
 XX Grandi G, Massignani V, Pizza M, Rappuoli R, Scarlato V;  
 WPI; 1999-327407/27.  
 DR P-PSDB; AAY38653.  
 XX Proteins from Neisseria meningitidis and N. gonorrhoeae useful for  
 PT diagnosis, treatment and prevention of infection  
 XX  
 PS Claim 9; Page 208; 524pp; English.  
 XX Nucleotide sequences AAZ11972-721358 represent open reading frames  
 CC (ORFs) of Neisseria meningitidis and N. gonorrhoeae which encode  
 CC antigenic proteins (see AAY38499-Y38944). The antigenic proteins, their  
 CC fragments, their nucleic acids and antibodies are used for diagnosis,  
 CC prevention (as vaccines) or treatment of Neisseria infections,  
 CC such as meningitis, septicaemia and gonorrhea. Both organisms  
 CC are closely related. Fragments of the nucleic acids are useful  
 CC as hybridisation probes and antisense reagents.  
 XX  
 SQ Sequence 1575 BP; 385 A; 387 C; 392 G; 411 T; 0 other;

Query Match 31.6%; Score 60.6; DB 20; Length 1575;  
 Best Local Similarity 84.2%; Pred. No. 1.7e-10;  
 Matches 80; Conservative 0; Mismatches 14; Indels 1; Gaps 1;  
 QY 98 TCATTGGGATAAACAAATATTCGCGCTTCGCGAATGTTCAAGCTAGCTGCATCA 157  
 Db 1575 TTATTGGGATAAAC-ATATTCGCGCTTCGCGAATGTTCAAGCTAGCTGCATCGC 1517

QY 158 CCGTAATCAGGTTGCCGTTACCGAGCCTTCGAGA 192  
 DB 1516 CCGTAATCAGGTTGCCGTTACCGAGCCTTCGCGA 1482

## RESULT 13

AA15323/C  
 ID AAA15323 standard; DNA; 1575 BP.

XX AC AAA15323;  
 XX 04-SEP-2000 (first entry)  
 XX DNA encoding a polypeptide of a *Neisseria* pathogenic strain.  
 DE Pathogenic strain; *Neisseria*; vaccine; *Neisseria* infection; ss.  
 XX *Neisseria* gonorrhoeae.  
 XX Key Location/Qualifiers  
 XX CDS 1..1575  
 XX /\*tag= a

XX WO200026375-A2.

PN 11-MAY-2000.

XX 28-OCT-1999; 99WO-FR02643.

XX 30-OCT-1998; 98FR-0013693.

XX (INMR ) PASTEUR MERIEUX SERUMS & VACCINS SA.  
 XX (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.

XX Aujane L, Bouchardon A, Renauld-Mongenie G, Rokhi B, Nassif X;  
 XX Tinsley C, Perrin A;  
 XX WPI; 2000-365622/31.  
 DR P-PSDB; AA93293.

XX New polypeptide specific for pathogenic *Neisseria* useful in therapeutic  
 PT or preventative vaccines and for diagnosis  
 XX Claim 4; Page 137-139; 187pp; French.

XX The present sequence encodes a protein that is specific for pathogenic  
 CC strains of *Neisseria*. The polynucleotides, polypeptides, or their  
 CC antigenic fragments, are used in vaccines to treat or protect against  
 CC *Neisseria* infections, particularly by *N. meningitidis*. The  
 CC polynucleotide sequence is also used for recombinant production of  
 CC the polypeptide and to produce attenuated *Neisseria* strains that  
 CC overexpress it, or express it in a non-toxic mutant form.

XX Sequence 1575 BP; 385 A; 387 C; 392 G; 411 T; 0 other;

Query Match 31.6%; Score 60.6; DB 21; Length 1575;  
 Best Local Similarity 84.2%; Pred. No. 1.7e-10;  
 Matches 80; Conservative 0; Mismatches 14; Indels 1; Gaps 1;

QY 98 TCATTGCGGATAACAATATTCGCCCTTCCGCGAATGTTCAAGCTAGCCTGCATCA 157  
 DB 1575 TTATTGCGGATAAAC-ATATTCGCCCTTCCGCGTGGAAATGTTCAAGCTGCCTGCATCGC 1517

QY 158 CCGTAATCAGGTTGCCGTTACCGAGCCTTCGAGA 192  
 DB 1516 CCGTAATCAGGTTGCCGTTACCGAGCCTTCGCGA 1482

## RESULT 14

AB241628/C  
 ID AB241628 standard; DNA; 1467 BP.

XX

AC AB241628;  
 XX 07-MAR-2003 (first entry)  
 XX N. gonorrhoeae nucleotide sequence SEQ ID 7845.  
 DE Antibacterial; infection; vaccine; gene therapy; gene; ds.  
 XX *Neisseria* gonorrhoeae.  
 OS WO200279243-A2.  
 PN 10-OCT-2002.  
 XX 12-FEB-2002; 2002WO-1B02069.  
 XX 12-FEB-2001; 2001GB-0003424.  
 XX (CHIR-) CHIRON SPA.  
 XX Fontana MR, Pizza M, Masignani V, Monaci E;  
 XX WPI; 2003-058415/05.  
 DR P-PSDB; ABP80658.  
 XX New protein from *Neisseria* gonorrhoeae, useful for the manufacture of a  
 PT medicament for treating or preventing *N. gonorrhoeae* infection -  
 XX Disclosure; Page 760-761; 815pp; English.

XX The present invention relates to proteins from *Neisseria* gonorrhoeae.  
 CC Also disclosed are the nucleic acid molecules encoding the proteins and  
 CC antibodies that specifically bind to the proteins. The composition  
 CC comprising the protein, nucleic acid or antibody is useful for the  
 CC manufacture of a medicament for treating or preventing *N. gonorrhoeae*  
 CC infection, this may be in the form of a vaccine or gene therapy.  
 CC Sequences given in records AB237706-AB242016 represent nucleic acid  
 CC molecules of the invention.

XX Sequence 1467 BP; 356 A; 358 C; 377 G; 376 T; 0 other;  
 Query Match 30.8%; Score 59.2; DB 25; Length 1467;  
 Best Local Similarity 84.8%; Pred. No. 5.1e-10;  
 Matches 78; Conservative 0; Mismatches 13; Indels 1; Gaps 1;

QY 101 TTGCGGATAACAATATTCGCCCTTCCGCGAATGTTCAAGCTAGCCTGCATCACCG 160  
 DB 1467 TTGCGGATAAAC-ATATTCGCCCTTCCGCGTGGAAATGTTCAAGCTGCCTGCATCCCG 1409

QY 161 TAATCAGGTTGCCGTTACCGAGCCTTCGAGA 192

DB 1408 TAATCAGGTTGCCGTTACCGAGCCTTCGAGA 1377

## RESULT 15

AA40550/C

ID AA40550 standard; DNA; 12100 BP.

XX AC AA40550;

XX 17-DEC-2001 (first entry)

XX DNA encoding human prostate cancer antigen, Seq ID No 702.  
 XX Human; prostate cancer antigen; cytostatic; uropathic; diagnostic;  
 KW reproductive system; chromosomal marker; forensic; urinary disorder;  
 KW chronic nephritis; blood-related disorder; thrombosis; ds.

XX Homo sapiens.

XX WO200155316-A2.

XX 02-AUG-2001.





Isolated polypeptide for treating, preventing and/or prognosing disorders related to the reproductive system including prostate cancer and also for testing and detection e.g. diagnosis -

Disclosure; SEQ ID No 702; 546pp; English.

The invention relates to novel isolated human prostate cancer antigen polynucleotides' (I) and polypeptides (II). (I) and (II) are useful for preventing, treating or ameliorating a medical condition when administered. (I), (II) and the antibody to (II) are useful for treating, preventing and/or prognosing disorders related to the reproductive system including prostate cancers; urinary disorders e.g. chronic nephritis; and blood-related disorders e.g. thrombosis. (II) can be used for testing and detection e.g. as a chromosomal marker and in forensics. (I) and the anti-(I) antibody can be used in testing and detection in immunoassays. AA840061-AA840075 represent the human prostate cancer antigen coding sequences, and related PCR primers and sequences of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at: ftp.wipo.int/pub/published\_pct\_sequences.

Sequence 12100 BP; 3692 A; 2710 C; 2748 G; 2360 T; 0 other;

Query Match 18.4%; Score 35.4; DB 22; Length 12100;  
Best Local Similarity 51.6%; Pred. No. 0.18;  
Matches 81; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 1 GATCGCGTACTGGTTTTCATATTITGCAATAGCTTGTCGGTGGGCATCTCCCGA 60  
| | | | |  
Db 4343 GCTGTAGTAGTACTGGTGTGTTTTTTTTTCTCTGGGGTCAGTGGTAACATTCCCTT 4284  
| | | | |  
QY 61 CATCATCTAAATTTGTCCTTTATGGTTTTACGCCACTCATTGGCGATAACAATATCCC 120  
| | | | |  
Db 4283 CATCAATTTCTAATTTTCTTTATTTGAATCTTCTCATCTCTTTTATTAATCTAGTAGT 4224  
| | | | |  
QY 121 GCCTTGCCGTCCGGAATGTTCAAGCTAGCGCTGCATCA 157  
| | | | |  
Db 4223 GCCTATCTTAGTAGTAATTTTTCAAAAAACCTTGATTA 4187  
| | | | |

Search completed: November 15, 2003, 00:35:20  
Job time : 125.917 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 14, 2003, 23:58:39 ; Search time 1280.85 Seconds  
(without alignments)  
3643.257 Million cell updates/sec

Title: US-09-928-457-67

Perfect score: 192

Sequence: 1 GATCCGGTACTGGTGTTTT.....CGGTACCGAGCCTTCGAGA 192

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152338056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_nam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gsal:*
29: gb_gss2:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	36.6	19.1	281	10	BF870084 IL3-ETO11
C 2	35.4	18.4	1117	9	AL575232 AL575232
C 3	34.2	17.8	429	28	A2499482
C 4	34	17.7	710	28	BH975720

#### TITLE

Shotgun sequencing of the human transcriptome with ORF expressed

#### JOURNAL

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

#### MEDLINE

20202663

#### PUBMED

10737800

#### COMMENT

Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

#### ALIGNMENTS

RESULT 1  
BF870084/c BF870084 281 bp mRNA linear EST 17-JAN-2001  
LOCUS IL3-ETO116-281000-309-B12 ET0116 Homo sapiens cDNA, mRNA sequence.  
DEFINITION BF870084  
ACCESSION BF870084  
VERSION BF870084.1 GI:12260214  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 281)  
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,  
Ngai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Cosca, F.F.,  
Goldman, G.H., Carvalho, A.F., Matsumura, A., Baia, G.S., Simpson, D.H.,  
Grundstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,  
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
Simpson, A.J.G.

Shotgun sequencing of the human transcriptome with ORF expressed

#### JOURNAL

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

#### MEDLINE

20202663

#### PUBMED

10737800

#### COMMENT

Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,



musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (GI:4732114) [GB|AF129072.1], a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

```

BASE COUNT      141 a      72 c      95 g      121 t
ORIGIN
Query Match      17.8%; Score 34.2; DB 28; Length 429;
Best Local Similarity 62.1%; Pred. No. 1.e+02;
Matches 54; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY      32  TAGTCTGTGCGTGGGCATCTCCCGACATCATCTAAATTTGCTTTTATGGTTTAA 91
        |||||
Db      162  TCGTCTTTTGTCCAGGAGTTTCCATGACATATCTATGTTGGCTAGACTGCTTTA 221
        |||||

QY      92  CGCCATCATTTGGGATAAACAATAT 118
        |||||
Db      222  TACCACCACATGATACATACACAGT 248
        |||||

```

```

RESULT 4
BH975720
LOCUS      BH975720      710 bp      DNA      linear      GSS 02-OCT-2002
DEFINITION      odh5a02.b1 B.oleracea002 Brassica oleracea genomic, genomic survey
sequence.
ACCESSION      BH975720
VERSION      BH975720.1 GI:23458723
KEYWORDS      GSS.
SOURCE      Brassica oleracea
ORGANISM      Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids 1; Brassicales; Brassicaceae; Brassica.
REFERENCE      1 (bases 1 to 710)
AUTHORS      Delebaunty,K., Fewell,G., Fulton,L., McCombie,W.R., Miner,T., Nash
,W., Rabinowicz,P.D. and Wilson,R.K.
TITLE      Whole genome shotgun reads from Brassica oleracea
JOURNAL      Unpublished
COMMENT      Contact: Richard K. Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Plate: odh5 row: a column: 02
Seq primer: -21UPPOT forward
Class: shotgun
High quality sequence start: 12
High quality sequence stop: 551.
Location/Qualifiers
1..710
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/db_xref="taxon:3712"
/clone_lib="B.oleracea002"
/notes="Vector: POTw13; Whole genome shotgun library from
flowering buds. DNA was purified from a crude nuclear
prep using Brassica oleracea Tol000DH3 buds provided by
Thomas Osborn at the University of Wisconsin. Genomic
DNA was provided by Pablo Rabinowicz (CSHL) and the
shotgun library prepared at Washington University Genome

```

```

FEATURES
source
1..710
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/db_xref="taxon:3712"
/clone_lib="B.oleracea002"
/notes="Vector: POTw13; Whole genome shotgun library from
flowering buds. DNA was purified from a crude nuclear
prep using Brassica oleracea Tol000DH3 buds provided by
Thomas Osborn at the University of Wisconsin. Genomic
DNA was provided by Pablo Rabinowicz (CSHL) and the
shotgun library prepared at Washington University Genome

```

```

BASE COUNT      219 a      102 c      104 g      285 t
ORIGIN
Query Match      17.7%; Score 34; DB 28; Length 710;
Best Local Similarity 57.5%; Pred. No. 1.1e+02;
Matches 61; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY      11  CTTGTTTTCATATTTGTCATAGTCTTGTGCGTGGGCATCTTCCCGACATCATCTAA 70
        |||||
Db      477  CTCAGGTGCGCATCTTTTGTTACTTCTTCTGATTTAAATCTTCCAGTACACTTCTGT 536
        |||||

QY      71  ATTTGCTTTATGTTTGTACGCCACTCATTTGGGATACATA 116
        |||||
Db      537  TTTTCTTTTGTGACGTCCTTCTGTTTGTATATAAAAA 582
        |||||

```

```

RESULT 5
AO621897
LOCUS      AO621897      659 bp      DNA      linear      GSS 16-JUN-1999
DEFINITION      HS_3107_B1.H07.MR.CIT.Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate=3107 Col=13 Row=P, genomic survey
sequence.
ACCESSION      AO621897
VERSION      AO621897.1 GI:5084289
KEYWORDS      GSS.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 659)
AUTHORS      Mahaitas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.

```

```

TITLE      Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL      Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
MEDLINE      99380589
PUBMED      1049764
COMMENT      Contact: Mahaitas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end Web Server: http://www.htsc.washington.edu
Plate: 3107 row: P column: 13
Seq primer: M13 Reverse
Class: BAC ends
High quality sequence stop: 659.
Location/Qualifiers
1..659
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="Plate=3107 Col=13 Row=P"
/sex="male"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/notes="Organ: sperm; Vector: pBelBAC11; BAC clones in
E-Coli DH10B"

```

```

BASE COUNT      104 a      136 c      69 g      350 t
ORIGIN
Query Match      17.6%; Score 33.8; DB 28; Length 659;
Best Local Similarity 64.9%; Pred. No. 1.1e+02;
Matches 50; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY      16  TTTTTCATATTTGTCATAGTCTTGTGCGTGGGCATCTTCCCGACATCATCTAAATTG 75
        |||||
Db      495  TTTTTCCTGTCGCTTAACCTTCTTGTGTTGGGCTTATCCCTCCCTTCCCTTTAAATTT 554
        |||||

```



```

source
1. 342
/organism="Danio rerio"
/mol_type="mRNA"
/db_xref="taxon:7955"
/clone="ENORA091ZAL1"
/tissue_type="inner ear"
/dev_stage="embryonic"
/clone_lib="Danio rerio embryonic inner ear subtracted
cDNA"
BASE COUNT      79 a   94 c   75 g   94 t
ORIGIN
Query Match      17.1%; Score 32.8; DB 9; Length 342;
Best Local Similarity 49.4%; Pred.No.2.4e+02;
Matches 85; Conservative 0; Mismatches 87; Indels 0; Gaps 0;
QY      2 ATCCGGTACTTGGTTTTTCATATTTTGCATATGCTTGTGCGTGGCGCATCTTCCCGAC 61
DB      13 ACCGGGCGAGGTACTCTGCAATGATTTTTCAGCAGGTGATTTCCGAACGGCTCACAAATGA 72
QY      62 ATCATCTAAATTTGTCCTTTATTGCTTTTTTACGCCACTCATTTGGGGATAACAATATTCOG 121
DB      73 TTCTGATCAGTGTGTTATCATCGGTTCCCAACCCCTCATAGCATGTGTGGAGCTTTTCOG 132
QY      122 CTTTGGCGTCCGAATGTTCAACTAGCTGCATCCGTAATCAGGTGCC 173
DB      133 CAAAGTAGGCAGGTTTATTCCAGACGACCTTACCAGAGTCATCAGGCAGTC 184

RESULT 10
BO792412      841 bp mRNA linear EST 30-JUL-2002
LOCUS      EST 8867 veraison Grape berries Lambda Triplex2 Library Vitis
DEFINITION      vitifera cDNA clone CT006G02 3', mRNA sequence.
ACCESSION      BO792412
VERSION      BO792412.1 GI:22007378
KEYWORDS      EST.
SOURCE      Vitis vinifera
ORGANISM      Vitis vinifera
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; Vitaceae; Vitis.
REFERENCE      1 (bases 1 to 841)
AUTHORS      Abbal,P., Agasse,A., Ageorges,A., Atanassova,R., Barrieu,F.,
      Couture,C., Dedalechamp,F., Delrot,S., Glissant,D., Gimplet,J.,
      Hamdi,S., Romieu,C. and Terrier,N.
      Generation of Expressed Sequence Tag from Grape Berry (skin, pulp
      or seeds) at Various Developmental Stages
      Unpublished
      Contact: Hamdi S.
      UMR 619 - Equipe Biologie de la Vigne
      Universite de Bordeaux I, Institut National de la Recherche
      Agronomique
      71, Avenue Edouard Bourleaux, BP 81, 33883 Villenave D'Ornon Cedex,
      France
      Tel: 00-33-(0)5-57-12-25-50
      Fax: 00-33-(0)5-57-12-25-48
      Email: s.hamdi@bordeaux.inra.fr
      Seq primer: T7
      Location/Qualifiers
      1. 841
      /organism="Vitis vinifera"
      /mol_type="mRNA"
      /cultivat="Cabernet Sauvignon"
      /db_xref="taxon:29760"
      /clone="CT006G02"
      /dev_stage="veraison stage"
      /clone_lib="veraison Grape berries Lambda Triplex2
      Library"
      /note="Organ: Fruit without seeds; Vector: Lambda Triplex2
      ; Site 1: SfiIA; Site 2: SfiIB; Oriented library"
BASE COUNT      157 a   225 c   116 g   343 t

```





Email: enta@tigr.org  
Library was provided by David Severson  
Seq primer: T7  
Class: BAC ends.  
Location/Qualifiers  
source  
1. .863  
/organism="Aedes aegypti"  
/mol\_type="genomic DNA"  
/strain="Liverpool"  
/db\_xref="taxon:7159"  
/clone="NDL.27P10"  
/clone\_lib="Nore Dame Liverpool"  
/notes="Vector: pEBAC1; Site 1: Hind III; The library was prepared from whole body tissue of newly hatched L1 larvae by David Severson at the University of Notre Dame and Hongbin Zhang"  
BASE COUNT 219 a 174 c 162 g 308 t  
ORIGIN  
Query Match 16.9%; Score 32.4; DB 29; Length 863;  
Best Local Similarity 54.1%; Pred. NO. 2.8e+02;  
Matches 66; Conservative 0; Mismatches 56; Indels 0; Gaps 0;  
Qy 16 TTTTCATATTTGCATGCTTTGTCGGGCGCATCTTCCCGACATCATCTAAATTG 75  
Db 318 TTGTACTAGTTGTCCTCACTGATGTTGTGAACATTTCTCTCGTTCCTTCAGTAT 377  
Qy 76 TCATTATGTTTTCAGGCACCTCATTCGGGATAACAATATTCGCTTCGCTCGCGA 135  
Db 378 CCTCATTCGATCGTTTTCATCGATGTTGCTGAACACTTCTTCGTTTTCCTTCAGT 437  
Qy 136 AT 137  
Db 438 AT 439  
RESULT 14  
B10635/c  
LOCUS  
DEFINITION  
Arabisopsis thaliana (thale cress)  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids  
; eurosids. II; Brassicales; Brassicaceae; Arabidopsi  
Feng,J.J., Dewar,K., Buehler,E., Kim,C., Li,Y., Shinn,P., Sun,H. and Ecker,J.  
BAC End Sequences at ATGC  
Other\_GSSs: F18123-Sp6  
Contact: Ecker J.  
Arabidopsis Thaliana Genome Center  
University of Pennsylvania  
Dept. of Biology, University of Pennsylvania, Philadelphia, PA 19104  
Tel: 215-998-9384  
Fax: 215-998-8780  
Email: jacker@atgenome.bio.upenn.edu  
Seq primer: T7  
Class: BAC ends  
High quality sequence start: 154  
High quality sequence stop: 352.  
Features  
source  
1. .936  
/organism="Arabidopsis thaliana"  
/mol\_type="genomic DNA"  
/strain="Columbia"  
/db\_xref="taxon:3702"  
Email: enta@tigr.org  
Library was provided by David Severson  
Seq primer: T7  
Class: BAC ends.  
Location/Qualifiers  
source  
1. .863  
/organism="Aedes aegypti"  
/mol\_type="genomic DNA"  
/strain="Liverpool"  
/db\_xref="taxon:7159"  
/clone="NDL.27P10"  
/clone\_lib="Nore Dame Liverpool"  
/notes="Vector: pEBAC1; Site 1: Hind III; The library was prepared from whole body tissue of newly hatched L1 larvae by David Severson at the University of Notre Dame and Hongbin Zhang"  
BASE COUNT 219 a 174 c 162 g 308 t  
ORIGIN  
Query Match 16.9%; Score 32.4; DB 29; Length 863;  
Best Local Similarity 54.1%; Pred. NO. 2.8e+02;  
Matches 66; Conservative 0; Mismatches 56; Indels 0; Gaps 0;  
Qy 16 TTTTCATATTTGCATGCTTTGTCGGGCGCATCTTCCCGACATCATCTAAATTG 75  
Db 318 TTGTACTAGTTGTCCTCACTGATGTTGTGAACATTTCTCTCGTTCCTTCAGTAT 377  
Qy 76 TCATTATGTTTTCAGGCACCTCATTCGGGATAACAATATTCGCTTCGCTCGCGA 135  
Db 378 CCTCATTCGATCGTTTTCATCGATGTTGCTGAACACTTCTTCGTTTTCCTTCAGT 437  
Qy 136 AT 137  
Db 438 AT 439  
RESULT 14  
B10635/c  
LOCUS  
DEFINITION  
Arabisopsis thaliana (thale cress)  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids  
; eurosids. II; Brassicales; Brassicaceae; Arabidopsi  
Feng,J.J., Dewar,K., Buehler,E., Kim,C., Li,Y., Shinn,P., Sun,H. and Ecker,J.  
BAC End Sequences at ATGC  
Other\_GSSs: F18123-Sp6  
Contact: Ecker J.  
Arabidopsis Thaliana Genome Center  
University of Pennsylvania  
Dept. of Biology, University of Pennsylvania, Philadelphia, PA 19104  
Tel: 215-998-9384  
Fax: 215-998-8780  
Email: jacker@atgenome.bio.upenn.edu  
Seq primer: T7  
Class: BAC ends  
High quality sequence start: 154  
High quality sequence stop: 352.  
Features  
source  
1. .936  
/organism="Arabidopsis thaliana"  
/mol\_type="genomic DNA"  
/strain="Columbia"  
/db\_xref="taxon:3702"

Email: enta@tigr.org  
Library was provided by David Severson  
Seq primer: T7  
Class: BAC ends.  
Location/Qualifiers  
source  
1. .863  
/organism="Aedes aegypti"  
/mol\_type="genomic DNA"  
/strain="Liverpool"  
/db\_xref="taxon:7159"  
/clone="NDL.27P10"  
/clone\_lib="Nore Dame Liverpool"  
/note="Vector: pEBAC1; Site 1: Hind III; The library was prepared from whole body tissue of newly hatched L1 larvae by David Severson at the University of Notre Dame and Hongbin Zhang"  
BASE COUNT 219 a 174 c 162 g 308 t  
ORIGIN  
Query Match 16.9%; Score 32.4; DB 29; Length 863;  
Best Local Similarity 54.1%; Pred. NO. 2.8e+02;  
Matches 66; Conservative 0; Mismatches 56; Indels 0; Gaps 0;  
Qy 16 TTTTCATATTTGCATGCTTTGTCGGGCGCATCTTCCCGACATCATCTAAATTG 75  
Db 318 TTGTACTAGTTGTCCTCACTGATGTTGTGAACATTTCTCTCGTTCCTTCAGTAT 377  
Qy 76 TCATTATGTTTTCAGGCACCTCATTCGGGATAACAATATTCGCTTCGCTCGCGA 135  
Db 378 CCTCATTCGATCGTTTTCATCGATGTTGCTGAACACTTCTTCGTTTTCCTTCAGT 437  
Qy 136 AT 137  
Db 438 AT 439  
RESULT 14  
B10635/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids  
; eurosids. II; Brassicales; Brassicaceae; Arabidopsis.  
REFERENCE  
AUTHORS  
Feng, J.J., Dewar, K., Buehler, E., Kim, C., Li, Y., Shinn, P., Sun, H. and Ecker, J.  
TITLE  
BAC End Sequences at ATGC  
JOURNAL  
COMMENT  
Other\_GSSs: F18123-Sp6  
Contact: Ecker J.  
Arabidopsis Thaliana Genome Center  
University of Pennsylvania  
Dept. of Biology, University of Pennsylvania, Philadelphia, PA 19104  
Tel: 215-998-9384  
Fax: 215-998-8780  
Email: jacker@atgenome.bio.upenn.edu  
Seq primer: T7  
Class: BAC ends  
High quality sequence start: 154  
High quality sequence stop: 352.  
Features  
source  
1. .936  
/organism="Arabidopsis thaliana"  
/mol\_type="genomic DNA"  
/strain="Columbia"  
/db\_xref="taxon:3702"  
Email: enta@tigr.org  
Library was provided by David Severson  
Seq primer: T7  
Class: BAC ends.  
Location/Qualifiers  
source  
1. .863  
/organism="Aedes aegypti"  
/mol\_type="genomic DNA"  
/strain="Liverpool"  
/db\_xref="taxon:7159"  
/clone="NDL.27P10"  
/clone\_lib="Nore Dame Liverpool"  
/note="Vector: pEBAC1; Site 1: Hind III; The library was prepared from whole body tissue of newly hatched L1 larvae by David Severson at the University of Notre Dame and Hongbin Zhang"  
BASE COUNT 219 a 174 c 162 g 308 t  
ORIGIN  
Query Match 16.9%; Score 32.4; DB 29; Length 863;  
Best Local Similarity 54.1%; Pred. NO. 2.8e+02;  
Matches 66; Conservative 0; Mismatches 56; Indels 0; Gaps 0;  
Qy 16 TTTTCATATTTGCATGCTTTGTCGGGCGCATCTTCCCGACATCATCTAAATTG 75  
Db 318 TTGTACTAGTTGTCCTCACTGATGTTGTGAACATTTCTCTCGTTCCTTCAGTAT 377  
Qy 76 TCATTATGTTTTCAGGCACCTCATTCGGGATAACAATATTCGCTTCGCTCGCGA 135  
Db 378 CCTCATTCGATCGTTTTCATCGATGTTGCTGAACACTTCTTCGTTTTCCTTCAGT 437  
Qy 136 AT 137  
Db 438 AT 439  
RESULT 14  
B10635/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids  
; eurosids. II; Brassicales; Brassicaceae; Arabidopsis.  
REFERENCE  
AUTHORS  
Feng, J.J., Dewar, K., Buehler, E., Kim, C., Li, Y., Shinn, P., Sun, H. and Ecker, J.  
TITLE  
BAC End Sequences at ATGC  
JOURNAL  
COMMENT  
Other\_GSSs: F18123-Sp6  
Contact: Ecker J.  
Arabidopsis Thaliana Genome Center  
University of Pennsylvania  
Dept. of Biology, University of Pennsylvania, Philadelphia, PA 19104  
Tel: 215-998-9384  
Fax: 215-998-8780  
Email: jacker@atgenome.bio.upenn.edu  
Seq primer: T7  
Class: BAC ends  
High quality sequence start: 154  
High quality sequence stop: 352.  
Features  
source  
1. .936  
/organism="Arabidopsis thaliana"  
/mol\_type="genomic DNA"  
/strain="Columbia"  
/db\_xref="taxon:3702"



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 15, 2003, 00:06:39 ; Search time 29.9056 Seconds  
(without alignments)  
2833.774 Million cell updates/sec

Title: US-09-928-457-67  
Perfect score: 192  
Sequence: 1 GATCCGCGTACTTGCTTTT.....CCGTACGACCTTCGAGA 192

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA: \*  
1: /cgn2\_6/ptodata/2/ina/5A COMB.seq: \*  
2: /cgn2\_6/ptodata/2/ina/5B COMB.seq: \*  
3: /cgn2\_6/ptodata/2/ina/6A COMB.seq: \*  
4: /cgn2\_6/ptodata/2/ina/6B COMB.seq: \*  
5: /cgn2\_6/ptodata/2/ina/PTCUTUS COMB.seq: \*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	29.2	15.2	2868	4	US-08-780-562-8
2	29.2	15.2	3097	2	US-08-599-455B-1
3	29.2	15.2	3097	3	US-09-069-781B-1
4	29.2	15.2	3097	4	US-09-137-132-1
5	29.2	15.2	3097	4	US-08-864-564A-1
6	29.2	15.2	3097	4	US-09-094-410-1
7	29.2	15.2	3097	4	US-08-708-123D-1
8	29.2	15.2	3097	4	US-08-583-153A-1
9	29.2	15.2	3097	4	US-08-570-142D-1
10	29.2	15.2	3097	4	US-08-638-524B-1
11	29.2	15.2	3495	3	US-08-827-962-17
12	29.2	15.2	3650	3	US-08-837-635-5
13	29.2	15.2	3650	3	US-08-837-635-7
14	29.2	15.2	3650	3	US-08-827-962-16
15	29.2	15.2	3650	3	US-08-803-345-2
16	29.2	15.2	3854	2	US-08-599-455B-42
17	29.2	15.2	3854	3	US-09-069-781B-42
18	29.2	15.2	3854	4	US-09-137-132-42
19	29.2	15.2	3854	4	US-08-864-564A-42
20	29.2	15.2	3854	4	US-09-094-410-42
21	29.2	15.2	3854	4	US-08-708-123D-42
22	29.2	15.2	3854	4	US-08-638-524B-42
C 23	29.2	15.2	3255	2	US-09-047-026A-3
C 24	29.2	15.2	6854	1	US-08-468-036-4
C 25	29.2	15.2	6854	2	US-08-376-843-4
26	28.8	15.0	7218	1	US-08-232-463-14
C 27	28.8	15.0	11887	4	US-08-961-527-146

Sequence 100, Appl  
Sequence 1, Appl  
Sequence 77, Appl  
Sequence 565, Appl  
Sequence 519, Appl  
Sequence 493, Appl  
Sequence 609, Appl  
Sequence 695, Appl  
Sequence 2, Appl  
Sequence 1, Appl  
Sequence 22, Appl  
Sequence 1, Appl  
Sequence 17, Appl  
Sequence 35, Appl  
Sequence 151, Appl  
Sequence 151, Appl  
Sequence 5, Appl  
Sequence 4, Appl

ALIGNMENTS

RESULT 1  
US-08-780-562-8  
; Sequence 8, Application US/08780562  
; Patent No. 6541604  
; GENERAL INFORMATION:  
; APPLICANT: Matthews, William  
; APPLICANT: Bennett, Brian  
; TITLE OF INVENTION: WSX RECEPTOR  
; NUMBER OF SEQUENCES: 45  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 460 Point San Bruno Blvd  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Winpatin (Genentech)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/780,562  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/585005  
; FILING DATE: 01/08/97  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/  
; FILING DATE: 01/08/97  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lee, Wendy M.  
; REGISTRATION NUMBER: 40,378  
; REFERENCE/DOCKET NUMBER: P0986R1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415/225-1994  
; TELEFAX: 415/952-9881  
; TELEX: 910/371-7168  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2868 base pairs  
; TYPE: Nucleic Acid  
; STRANDEDNESS: Single  
; TOPOLOGY: Linear  
US-08-780-562-8

Query Match 15.2%; Score 29.2; DB 4; Length 2868;  
Best Local Similarity 59.8%; Pred. No. 2.1;  
Matches 49; Conservative 0; Mismatches 33; Indels 0; Gaps 0;



```

; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 61...2742
US-09-137-132-1

Query Match      15.2%; Score 29.2; DB 4; Length 3097;
Best Local Similarity 59.8%; Pred.No.2.1;
Matches 49; Conservative 0; Mismatches 33; Indels 0; Gaps 0

QY      65 ATCTAAATTCGTCITTTATTTGTTTACGGCAGTCATTCGGGATAACAATATTCGGCT 124
Db      267 ACCTAAATTTAATTCAGTGTGATATCTACGTTCTTGAGTTATCAAAACAGTCTTTCAC 326
QY      125 TCGCGTCGCGAATGTTCAAGCT 146
Db      327 TTGCTTTGGGAATGAGCAAGCT 348

RESULT 5
US-08-864-564A-1
; Sequence 1, Application US/08864564A
; Patent No. 6395498
; GENERAL INFORMATION:
; APPLICANT: Tartaglia, Louis A.
; APPLICANT: Tepper, Robert I.
; APPLICANT: Culpepper, Janice A.
; APPLICANT: White, David W.
; TITLE OF INVENTION: THE OB RECEPTOR AND METHODS FOR
; TITLE OF INVENTION: THE DIAGNOSIS AND TREATMENT OF BODY WEIGHT DISORDERS
; TITLE OF INVENTION: INCLUDING OBESITY AND CACHEXIA
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/864,564A
; FILING DATE: 28-MAY-1997
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 08/708,123
; FILING DATE: 03-SEP-1996
; APPLICATION NUMBER: 08/638,524
; FILING DATE: 26-APR-1996
; APPLICATION NUMBER: 08/599,455
; FILING DATE: 22-JAN-1996
; APPLICATION NUMBER: 08/583,153
; FILING DATE: 28-DEC-1995
; APPLICATION NUMBER: 08/570,142
; FILING DATE: 11-DEC-1995
; APPLICATION NUMBER: 08/569,485
; FILING DATE: 08-DEC-1995
; APPLICATION NUMBER: 08/566,622
; FILING DATE: 04-DEC-1995
; APPLICATION NUMBER: 08/562,663
; FILING DATE: 27-NOV-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Meiklejohn, Ph.D., Anita L.
; REGISTRATION NUMBER: 35,283
; REFERENCE/DOCKET NUMBER: 07334/019002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-5070
; TELEFAX: 617-542-8906

```

```
TELEX: 200154
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3097 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 61...2742
US-08-864-564A-1

Query Match 15.2%; Score 29.2; DB 4; Length 3097;
Best Local Similarity 59.8%; Pred. No. 2.1;
Matches 49; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 65 ATCTAAATTGCTTTATTGGTTTACGCCACTCATTCGGGATAAACAATATTCGGCT 124
Db 267 AGCTAAATTTAATTCAAGTGGTATCTACGTTCTCGAGTATCCAAACAGCTTCCACTG 326

QY 125 TGCCGTCGCGAATGTTCAAGCT 146
Db 327 TTGCTTTGGGAATGACGAAGT 348

RESULT 6
US-09-094-410-1
; Sequence 1, Application US/09094410
; Patent No. 648352
; GENERAL INFORMATION:
; APPLICANT: Tartaglia, Louis A.
; APPLICANT: Tepper, Robert I.
; APPLICANT: Culpepper, Janice A.
; APPLICANT: White, David W.
; TITLE OF INVENTION: THE OB RECEPTOR AND METHODS FOR
; TITLE OF INVENTION: THE DIAGNOSIS AND TREATMENT OF BODY WEIGHT DISORDERS,
; TITLE OF INVENTION: INCLUDING OBESITY AND CACHEXIA
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/094,410
; FILING DATE: 09-JUN-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/864,564
; FILING DATE: 28-MAY-1997
; APPLICATION NUMBER: 08/708,123
; FILING DATE: 03-SEP-1996
; APPLICATION NUMBER: 08/638,524
; FILING DATE: 26-APR-1996
; APPLICATION NUMBER: 08/599,455
; FILING DATE: 22-JAN-1996
; APPLICATION NUMBER: 08/593,153
; FILING DATE: 28-DEC-1995
; APPLICATION NUMBER: 08/570,142
; FILING DATE: 11-DEC-1995
; APPLICATION NUMBER: 08/569,485
; FILING DATE: 08-DEC-1995
; APPLICATION NUMBER: 08/566,622
; FILING DATE: 04-DEC-1995
; APPLICATION NUMBER: 08/562,663
; FILING DATE: 27-NOV-1995
```

```
ATTORNEY/AGENT INFORMATION:
; NAME: Meiklejohn, Ph.D., Anita L.
; REGISTRATION NUMBER: 35,283
; REFERENCE/DOCKET NUMBER: 07334/019003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-5070
; TELEFAX: 617-542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3097 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 61...2742
US-09-094-410-1

Query Match 15.2%; Score 29.2; DB 4; Length 3097;
Best Local Similarity 59.8%; Pred. No. 2.1;
Matches 49; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 65 ATCTAAATTGCTTTATTGGTTTACGCCACTCATTCGGGATAAACAATATTCGGCT 124
Db 267 AGCTAAATTTAATTCAAGTGGTATCTACGTTCTCGAGTATCCAAACAGCTTCCACTG 326

QY 125 TGCCGTCGCGAATGTTCAAGCT 146
Db 327 TTGCTTTGGGAATGACGAAGT 348

RESULT 7
US-08-708-123D-1
; Sequence 1, Application US/08708123D
; Patent No. 6482927
; GENERAL INFORMATION:
; APPLICANT: Tartaglia, Louis A.
; APPLICANT: Tepper, Robert I.
; APPLICANT: Culpepper, Janice A.
; APPLICANT: White, David W.
; TITLE OF INVENTION: THE OB RECEPTOR AND METHODS FOR
; TITLE OF INVENTION: THE DIAGNOSIS AND TREATMENT OF BODY WEIGHT DISORDERS,
; TITLE OF INVENTION: INCLUDING OBESITY AND CACHEXIA
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/708,123D
; FILING DATE: 03-SEP-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/638,524
; FILING DATE: 26-APR-1996
; APPLICATION NUMBER: 08/599,455
; FILING DATE: 22-JAN-1996
; APPLICATION NUMBER: 08/583,153
; FILING DATE: 28-DEC-1995
; APPLICATION NUMBER: 08/570,142
; FILING DATE: 11-DEC-1995
; APPLICATION NUMBER: 08/569,485
; FILING DATE: 04-DEC-1995
; APPLICATION NUMBER: 08/562,663
; FILING DATE: 27-NOV-1995
```

```

; NAME: Meiklejohn, Anita L.
; REGISTRATION NUMBER: 35,283
; REFERENCE/DOCKET NUMBER: 07334/016001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-5070
; TELEFAX: 617-542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3097 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 61...2742
; US-08-583-153A-1

Query Match 15.2%; Score 29.2; DB
Best Local Similarity 59.8%; Pred. No. 2.1;
Matches 49; Conservative 0; Mismatches

QY 65 ATCTAAATTTGCTTTATTGGTTTTCAGCCATCT
Db 267 AGCTAAATTTAATTCAGGTGATCTACGTTCTCTG
QY 125 TGCCTCGCGAATGTTCAAGCT 146
Db 327 TTGCTTTGGGAATGAGCAAGGT 348

RESULT 9
US-08-570-142D-1
; Sequence 1, Application US/08570142D
; Patent No. 6509189
; GENERAL INFORMATION:
; APPLICANT: Tartaglia, Louis A.
; APPLICANT: Tepper, Robert I.
; APPLICANT: Culpepper, Janice A.
; TITLE OF INVENTION: THE OB RECEPTOR AND MET
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT
; TITLE OF INVENTION: OBESITY AND CACHEXIA
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESS: Fish & Richardson, P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FASSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/570,142D
; FILING DATE: 11-DEC-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/569,485
; FILING DATE: 08-DEC-1995
; APPLICATION NUMBER: 08/566,622
; FILING DATE: 04-DEC-1995
; APPLICATION NUMBER: 08/562,663
; FILING DATE: 27-NOV-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Meiklejohn, Ph.D., Anita L.
; REGISTRATION NUMBER: 35,283
; REFERENCE/DOCKET NUMBER: 07334/014001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-5070
; TELEFAX: 617-542-8906

```

```
/
/ TELEX: 200154
/ INFORMATION FOR SEQ ID NO: 1:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 3097 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: unknown
/ MOLECULE TYPE: cDNA
/ FEATURE:
/ NAME/KEY: Coding Sequence
/ LOCATION: 61...2742
/ US-08-570-142D-1

Query Match      15.2%; Score 29.2; DB 4; Length 3097;
Best Local Similarity 59.8%; Pred. No. 2.1;
Matches 49; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 65 ATCTAAATTTGCTTTATTGGTTTTTACGCCACTCATTGGCGGATAAACAATATTCGGCT 124
Db 267 ACCTAAATTTAATTCAGTGGTATCTACGTTCTCGATTATCCAAACAGCTTCCACTG 326

QY 125 TGGCGTCGCGAATGTTCAAGCT 146
Db 327 TTGCTTTGGGAATGAGCAAGGT 348

RESULT 10
US-08-638-524B-1
/ Sequence 1, Application US/08638524B
/ Patent No. 6548269
/ GENERAL INFORMATION:
/ APPLICANT: Tartaglia, Louis A.
/ APPLICANT: Tepper, Robert I.
/ APPLICANT: Culpepper, Janice A.
/ APPLICANT: White, David W.
/ TITLE OF INVENTION: THE OB RECEPTOR AND METHODS FOR THE
/ TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF BODY WEIGHT DISORDERS, INCLUDING OB
/ TITLE OF INVENTION: CACHEXIA
/ NUMBER OF SEQUENCES: 50
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Fish & Richardson, P.C.
/ STREET: 225 Franklin Street
/ CITY: Boston
/ STATE: MA
/ COUNTRY: US
/ ZIP: 02110-2804
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette
/ COMPUTER: IBM Compatible
/ OPERATING SYSTEM: Windows95
/ SOFTWARE: FastSeq for Windows Version 2.0
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/638,524B
/ FILING DATE: 26-APR-1996
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 08/599,455
/ FILING DATE: 22-JAN-1996
/ APPLICATION NUMBER: 08/583,153
/ FILING DATE: 28-DEC-1995
/ APPLICATION NUMBER: 08/570,142
/ FILING DATE: 11-DEC-1995
/ APPLICATION NUMBER: 08/569,485
/ FILING DATE: 08-DEC-1995
/ APPLICATION NUMBER: 08/566,622
/ FILING DATE: 04-DEC-1995
/ APPLICATION NUMBER: 08/562,663
/ FILING DATE: 27-NOV-1995
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Meiklejohn Ph.D., Anita L.
/ REGISTRATION NUMBER: 35,283
/ REFERENCE/DOCKET NUMBER: 07334/018001
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 617-542-5070
```

```
/
/ TELEFAX: 617-542-8906
/ TELEX: 200154
/ INFORMATION FOR SEQ ID NO: 1:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 3097 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: unknown
/ MOLECULE TYPE: cDNA
/ FEATURE:
/ NAME/KEY: Coding Sequence
/ LOCATION: 61...2742
/ US-08-638-524B-1

Query Match      15.2%; Score 29.2; DB 4; Length 3097;
Best Local Similarity 59.8%; Pred. No. 2.1;
Matches 49; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 65 ATCTAAATTTGCTTTATTGGTTTTTACGCCACTCATTGGCGGATAAACAATATTCGGCT 124
Db 267 ACCTAAATTTAATTCAGTGGTATCTACGTTCTCGATTATCCAAACAGCTTCCACTG 326

QY 125 TGGCGTCGCGAATGTTCAAGCT 146
Db 327 TTGCTTTGGGAATGAGCAAGGT 348

RESULT 11
US-08-827-962-17
/ Sequence 17, Application US/08827962A
/ Patent No. 6258944
/ GENERAL INFORMATION:
/ APPLICANT: VERCK & CO., INC.
/ TITLE OF INVENTION: OB RECEPTOR ISOFORMS AND NUCLEIC ACIDS
/ TITLE OF INVENTION: ENCODING THEM
/ FILE REFERENCE: 19693
/ CURRENT APPLICATION NUMBER: US/08/827,962A
/ CURRENT FILING DATE: 1997-05-06
/ PRIOR APPLICATION NUMBER: 60/016,899
/ PRIOR FILING DATE: 1996-05-06
/ NUMBER OF SEQ ID NOS: 21
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 17
/ LENGTH: 3495
/ TYPE: DNA
/ ORGANISM: Rattus No. 6258944vegicus
/ US-08-827-962-17

Query Match      15.2%; Score 29.2; DB 3; Length 3495;
Best Local Similarity 59.8%; Pred. No. 2.2;
Matches 49; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 65 ATCTAAATTTGCTTTATTGGTTTTTACGCCACTCATTGGCGGATAAACAATATTCGGCT 124
Db 281 ACCTAAATTTAATTCAGTGGTATCTACGTTCTCGATTATCCAAACAGCTTCCACTG 340

QY 125 TGGCGTCGCGAATGTTCAAGCT 146
Db 341 TTGCTTTGGGAATGAGCAAGGT 362

RESULT 12
US-08-837-635-5
/ Sequence 5, Application US/08837635
/ Patent No. 600798
/ GENERAL INFORMATION:
/ APPLICANT: ROSENBLUM, CHARLES. I.
/ APPLICANT: VAN DER PLOEG, LEONARDUS, H.T.
/ APPLICANT: QURESHI, SAJJAD, A.
/ APPLICANT: CULLY, DORIS, F.
/ APPLICANT: HESS, JOHN W.
/ APPLICANT: TOTA, MICHAEL, R.
/ APPLICANT: CHEN, FANG
```



```

; TITLE OF INVENTION: LEPTIN ASSAY
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: JOANNE M. GIESSEY - MERCK & CO., INC.
; STREET: 126 EAST LINCOLN AVENUE - P.O. BOX 2000
; CITY: RAHWAY
; STATE: NJ
; COUNTRY: USA
; ZIP: 07065
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/837,635
; FILING DATE: 21-APR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/016,051
; FILING DATE: 22-APR-1996
; FILING DATE: 15-NOV-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: GIESSEY, JOANNE M
; REGISTRATION NUMBER: 32,838
; REFERENCE/DOCKET NUMBER: 19686Y
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 732-594-3046
; TELEFAX: 732-594-4720
; TELEX:
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3650 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Other
; US-08-837-635-5

Query Match 15.2%; Score 29.2; DB 3; Length 3650;
Best Local Similarity 59.8%; Pred. No. 2.3;
Matches 49; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 65 ATCTAAATTTGCTTTTATTTAGCCACCTCATTCGGGATAAACAATATTCGGCT 124
Db 281 AGCTAAATTTAATTCACCTGTTCTAGCTTCTGAGTATCCAAACCAATTTCCACTG 340

QY 125 TCCGTCGCGAATGTTCAAGT 146
Db 341 TTGCTTTGGGAATGACCAAGT 362

RESULT 13
US-08-837-635-7
; Sequence 7, Application US/08837635
; Patent No. 6007998
; GENERAL INFORMATION:
; APPLICANT: ROSENBLUM, CHARLES. I.
; APPLICANT: VAN DER PLOEG, LEONARDUS, H.T.
; APPLICANT: QURESHI, SAJJAD, A.
; APPLICANT: CULLY, DORIS, F.
; APPLICANT: HESS, JOHN W.
; APPLICANT: TOTA, MICHAEL, R.
; APPLICANT: CHEN, FANG
; TITLE OF INVENTION: LEPTIN ASSAY
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: JOANNE M. GIESSEY - MERCK & CO., INC.
; STREET: 126 EAST LINCOLN AVENUE - P.O. BOX 2000
; CITY: RAHWAY
; STATE: NJ
; COUNTRY: USA

```

```

; ZIP: 07065
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/837,635
; FILING DATE: 21-APR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/016,051
; FILING DATE: 22-APR-1996
; FILING DATE: 15-NOV-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: GIESSEY, JOANNE M
; REGISTRATION NUMBER: 32,838
; REFERENCE/DOCKET NUMBER: 19686Y
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 732-594-3046
; TELEFAX: 732-594-4720
; TELEX:
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3650 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Other
; US-08-837-635-7

Query Match 15.2%; Score 29.2; DB 3; Length 3650;
Best Local Similarity 59.8%; Pred. No. 2.3;
Matches 49; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 65 ATCTAAATTTGCTTTTATTTAGCCACCTCATTCGGGATAAACAATATTCGGCT 124
Db 281 AGCTAAATTTAATTCACCTGTTCTAGCTTCTGAGTATCCAAACCAATTTCCACTG 340

QY 125 TCCGTCGCGAATGTTCAAGT 146
Db 341 TTGCTTTGGGAATGACCAAGT 362

RESULT 14
US-08-827-962-16
; Sequence 16, Application US/08827962A
; Patent No. 6258944
; GENERAL INFORMATION:
; APPLICANT: MERCK & CO., INC.
; TITLE OF INVENTION: OB RECEPTOR ISOFORMS AND NUCLEIC ACIDS
; TITLE OF INVENTION: ENCODING THEM
; FILE REFERENCE: 19693
; CURRENT APPLICATION NUMBER: US/08/827,962A
; CURRENT FILING DATE: 1997-05-06
; PRIOR APPLICATION NUMBER: 60/016,899
; PRIOR FILING DATE: 1996-05-06
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 3650
; TYPE: DNA
; ORGANISM: Rattus No. 6258944vegicus
; US-08-827-962-16

Query Match 15.2%; Score 29.2; DB 3; Length 3650;
Best Local Similarity 59.8%; Pred. No. 2.3;
Matches 49; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 65 ATCTAAATTTGCTTTTATTTAGCCACCTCATTCGGGATAAACAATATTCGGCT 124
Db 281 AGCTAAATTTAATTCACCTGTTCTAGCTTCTGAGTATCCAAACCAATTTCCACTG 340

```

QY 125 TGCCGTGCGGAATGTTCAAGCT 146  
 Db 341 TTGCTTTGGGAATGAGCAAGT 362

## RESULT 15

US-08-803-346-2  
 ; Sequence 2, Application US/08803346  
 ; Patent No. 6281346  
 ; GENERAL INFORMATION:  
 ; APPLICANT: HESS, JOHN W.  
 ; APPLICANT: CASKEY, C. THOMAS  
 ; APPLICANT: LIU, QINGYUN  
 ; APPLICANT: PHILLIPS, MICHAEL SEAN  
 ; TITLE OF INVENTION: RAT OB RECEPTORS AND NUCLEOTIDES  
 ; TITLE OF INVENTION: ENCODING THEM  
 ; NUMBER OF SEQUENCES: 77  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: JOANNE M. GIESSEY - MERCK & CO., INC.  
 ; STREET: 126 EAST LINCOLN AVENUE - P.O. BOX 2000  
 ; CITY: RAHWAY  
 ; STATE: NJ  
 ; COUNTRY: USA  
 ; ZIP: 07065  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: DOS  
 ; SOFTWARE: Fast-Seq for Windows Version 2.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/803,346  
 ; FILING DATE: 20-FEB-1997  
 ; CLASSIFICATION: 536  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER:  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: GIESSEY, JOANNE M.  
 ; REGISTRATION NUMBER: 32,838  
 ; REFERENCE/DOCKET NUMBER: 19642Y  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 732-594-3046  
 ; TELEFAX: 732-594-4720  
 ; TELEX:  
 ; INFORMATION FOR SEQ ID NO: 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 3650 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: cDNA  
 ; US-08-803-346-2

Query Match 15.2%; Score 29.2; DB 3; Length 3650;  
 Best Local Similarity 59.8%; Pred. No. 2.3;  
 Matches 49; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 65 ATCTAAATTTCTTTTATTTTTCAGCCACTTCATTCGGATAAACAATATTCGCCT 124  
 Db 281 AGCTAAATTTAATCACTGTAATCTAGTTTCTGAGTTATCCAAACCAATTTTCACTG 340  
 QY 125 TGCCGTGCGGAATGTTCAAGCT 146  
 Db 341 TTGCTTTGGGAATGAGCAAGT 362

Search completed: November 15, 2003, 08:08:52  
 Job time : 36.9056 secs





PRIOR APPLICATION NUMBER: 60/246,473	PRIOR FILING DATE: 2000-11-08
PRIOR APPLICATION NUMBER: 60/246,532	PRIOR FILING DATE: 2000-11-08
PRIOR APPLICATION NUMBER: 60/249,216	PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,210	PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/226,681	PRIOR FILING DATE: 2000-08-22
PRIOR APPLICATION NUMBER: 60/225,759	PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/225,213	PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/227,182	PRIOR FILING DATE: 2000-08-22
PRIOR APPLICATION NUMBER: 60/225,214	PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/235,833	PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: 60/230,438	PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/215,135	PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: 60/225,266	PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/249,218	PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,208	PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,213	PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,212	PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,207	PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,245	PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,244	PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,217	PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,211	PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,215	PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,264	PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,214	PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,297	PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/232,400	PRIOR FILING DATE: 2000-09-14
PRIOR APPLICATION NUMBER: 60/231,242	PRIOR FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: 60/232,081	PRIOR FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: 60/232,080	PRIOR FILING DATE: 2000-09-14
PRIOR APPLICATION NUMBER: 60/233,063	PRIOR FILING DATE: 2000-09-14
PRIOR APPLICATION NUMBER: 60/232,397	PRIOR FILING DATE: 2000-09-14
PRIOR APPLICATION NUMBER: 60/232,399	PRIOR FILING DATE: 2000-09-14
PRIOR APPLICATION NUMBER: 60/232,401	PRIOR FILING DATE: 2000-09-14

```

; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/241,808
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,826
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,786
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,221
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/246,475
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/231,243
; PRIOR FILING DATE: 2000-09-08

Query Match      18.4%; Score 35.4; DB 14; Length 12100;
Best Local Similarity 51.6%; Pred. No. 0.84;
Matches 81; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 1 GATCCGCTACTTGGTTTTCATATTTCATAGTCTTGTGCGTGGCGCATCTTCCCGA 60
Db 4343 GCTCGTAGTACTGTTGTTATTTTCTCTGGGTCACTGTTAAATCCCTT 4284
QY 61 CATCATCTAAATTCCTTTATTTGTTTTCAGCCACTCATTCGCGATPAAACAATATTC 120
Db 4283 CATCATTTCTAATTTCTTTATTTTGAATCTTCTCATCTCTTTATTAATCTAGTAGT 4224
QY 121 GCCTTCCGTCGCGAATGTTCAAGCTAGCTGCATCA 157
Db 4223 GCCCTATCTTAGTAATTTTCAAAAAACCTTGATTA 4187

RESULT 5
US-09-781-558-3/c
; Sequence 3, Application US/09781558
; Patent No. US20020034778A1
; GENERAL INFORMATION:
; APPLICANT: Beasley et al.
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000756
; CURRENT APPLICATION NUMBER: US/09/781,558
; PRIOR FILING DATE: 2001-02-13
; PRIOR APPLICATION NUMBER: US 60/199,811
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: US 09/641,426
; PRIOR FILING DATE: 2000-08-18
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 74586
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(74586)
; OTHER INFORMATION: n = A,T,C or G

US-09-781-558-3
; Sequence 3, Application US/10388650
; Publication No. US20030166151A1
; GENERAL INFORMATION:
; APPLICANT: BRANDON et al.
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000756-CIPCON
; CURRENT APPLICATION NUMBER: US/10/388,650
; CURRENT FILING DATE: 2003-03-17
; PRIOR APPLICATION NUMBER: US 60/199,811
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: US 09/641,426
; PRIOR FILING DATE: 2000-08-18
; PRIOR APPLICATION NUMBER: US 09/781,558
; PRIOR FILING DATE: 2001-02-13
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 74586
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(74586)
; OTHER INFORMATION: n = A,T,C or G

US-10-388-650-3
Query Match      17.1%; Score 32.8; DB 12; Length 74586;
Best Local Similarity 53.0%; Pred. No. 12;
Matches 70; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 15 GTTTTTCATATTTGCATAGTCTTGTGCGTGGCGATCTTCCCGACATCATCTAAATTT 74
Db 48079 GTACTTCATCTGTGCTAAACTTTTGGCGTATTGTAACCTTCTAAGTTGCTTCTCTTTT 48020
QY 75 GCTTTTATTTGTTTTCAGCCACTCATTCGCGATTAACAATATTCGCGTCCGCGCG 134
Db 48019 GTTTATGCTTTTCTTCCCTAATTCATTGAAGTCAACAATTTGACTCATTTACCTCTGA 47960
QY 135 AATGTTCAAGCT 146
Db 47959 ATTGCTCTACCT 47948

RESULT 7
US-10-108-605-348/c
; Sequence 348, Application US/10108605
; Publication No. US20020160934A1
; GENERAL INFORMATION:
; APPLICANT: Broadus, Julie
; APPLICANT: Stam, Lynn
; APPLICANT: Bachmann, Jane
; APPLICANT: Kamdar, Kim
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES FROM DROSOPHILA MELANOGASTER THAT ENCODE
; TITLE OF INVENTION: PROTEINS ESSENTIAL FOR LARVAL VIABILITY AND USES THEREOF
; FILE REFERENCE: 31133B
; CURRENT APPLICATION NUMBER: US/10/108,605
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: US 09/761,142
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/176,418
; PRIOR FILING DATE: 2000-01-14
; NUMBER OF SEQ ID NOS: 361
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 348
; LENGTH: 3891
; TYPE: DNA

```

```
; ORGANISM: Drosophila melanogaster
US-10-108-605-348

Query Match      16.5%; Score 31.6; DB 13; Length 3891;
Best Local Similarity 52.2%; Pred. No. 9.1;
Matches 70; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 9 TACTTGGTTTTCATATTTTGATAGTCTTTGTGCGTGGGATCTTCCCGACATCATCT 68
Db 1290 TTCTGCGGTGTTGCTGTATCACTTGTCTGTGCTGTGCGAGTTTCCCATCTCCGGT 1231

QY 69 AAATTGCTTTTATTTGGTTTTTACGCCACTATTCGGGATAAACAATATTCGGCTTCCC 128
Db 1230 AGACATGGCTTGAGCGGCGCTTCCTCCAGTCTTTCGATCCAGATCGGTCATAGTCCC 1171

QY 129 GTCGGAATGTTCA 142
Db 1170 ACCTTAATGGCA 1157

RESULT 8
US-10-090-002-3/c
; Sequence 3, Application US/10090002
; Publication No. US20030175926A1
; GENERAL INFORMATION:
; APPLICANT: YAN, Chunhua et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001316
; CURRENT APPLICATION NUMBER: US/10/090,002
; CURRENT FILING DATE: 2002-03-08
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 15400
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-090-002-3.

Query Match      16.1%; Score 31; DB 12; Length 15400;
Best Local Similarity 55.0%; Pred. No. 25;
Matches 61; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 81 ATTGTTTTAGCCACATTCGCGATAAACAATATTCGCGCTTGGCGTTCGCGAATGTT 140
Db 2589 AGTGATGCCCGCCGCCACCCCTGGGATGTCTGGGGTGTCTCTTCCCTCTCTCC 2530

QY 141 CAAGTAGCGCTCATCACCGTAATCAGGTGCGCGTTACCGAGCCTTCGAG 191
Db 2529 CCGACTCCCGCGCTCTCCGAATCAGGTTTCCAGCAGCTGGGCTTCCAG 2479

RESULT 9
US-09-974-300-6709/c
; Sequence 6709, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6709
; LENGTH: 197

; TYPE: DNA
; ORGANISM: Bacillus clausii
US-09-974-300-6709

Query Match      15.8%; Score 30; DB 10; Length 197;
Best Local Similarity 61.8%; Pred. No. 9;
Matches 48; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 38 TGTCGGTGGGCATCTTCCCGACATCATCTAAATTTGCTTTTATTGGTTTTTACGCCAC 97
Db 82 TTCTTCCGTTCTTTTGTCCGCTTAATGTAGATTTGGCTTTTCTTTTACGCAAC 23

QY 98 TCATTGGGATAAACAAT 115
Db 22 CAATTGCAGTCGCAGAT 5

RESULT 10
US-10-027-632-9875
; Sequence 9875, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9875
; LENGTH: 1136
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-9875

Query Match      15.5%; Score 29.8; DB 12; Length 1136;
Best Local Similarity 50.3%; Pred. No. 21;
Matches 73; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 3 TCCGCGTACTTGGTTTTTCATATTTTCATAGTCTTGTTCGGTGGGATCTTCCCCGACA 62
Db 874 TCCAGGTCTTGTTCCTCATTCGGAAACATGGTCTTGTGACGAGTAATCTTGAACCCG 933

QY 63 TCATCTAAATTTGCTTTTATTGTTTTTACGCCACATTCGCGATAAACAATATTCGCG 122
Db 934 TTCTTTAAATATTTCTTTTGGTCTATTTAAAGCCAGCGGAGGGGTGGAAGCCCTT 993

QY 123 CTTCGCGTCGGAATGTTCAAGCTA 147
Db 994 TATTGCTCAGGAAGTGTCATGTA 1018

RESULT 11
US-10-027-632-9875
; Sequence 9875, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
```

```

; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108927.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325920
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9875
; LENGTH: 1136
; TYPE: DNA
; ORGANISM: Human
; US-10-027-632-9875

```

Query Match	15.5%	Score 29.8;	DB 13;	Length 1136;
Best Local Similarity	50.3%;	Pred. No. 21;		
Matches	73;	Conservative 0;	Mismatches 72;	Indels 0; Gaps 0
QY	3	TCCGGCTACTGGTTTTTCATATTTTCATAGTCTTGTCGTCGGCGCATCTTCCCGGACA	62	
Db	874	TCCAGGTTCTTGTTCCTCATTTGGGAACATGGTCTTGTGACGATAATCTTTGAACCCCG	933	
QY	63	TCATCTAAATTTGCTTTTTATTGGTTTTTACGCCACTCATTTGCCGATAAACAAATATTCGC	122	
Db	934	TTTCTTAAATATCTTTTGGTCAATTTAAGCCAGCGGGAAGGGTGGAAAAGGCCCTT	993	
QY	123	CTTGCGTCGCGAATGTTCAAGCTA	147	
Db	994	TATTTGCTCAGGAAGGTGCATGTA	1018	

```

RESULT 12
US-10-277-161-70
; Sequence 70, Application US/10277161
; Publication No. US20030194696A1
; GENERAL INFORMATION:
; APPLICANT: Zauderer, Maurice
; APPLICANT: Smith, Ernest S.
; TITLE OF INVENTION: Methods of Producing a Library and Methods of Selecting Polynucle
; TITLE OF INVENTION: of Interest
; FILE REFERENCE: 1821.0050006
; CURRENT APPLICATION NUMBER: US/10/277,161
; CURRENT FILING DATE: 2002-10-22
; PRIOR APPLICATION NUMBER: 60/192,586
; PRIOR FILING DATE: 2000-03-28
; PRIOR APPLICATION NUMBER: 60/203,343
; PRIOR FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 60/263,226
; PRIOR FILING DATE: 2001-01-23
; PRIOR APPLICATION NUMBER: 60/271,426
; PRIOR FILING DATE: 2001-02-27
; PRIOR APPLICATION NUMBER: 09/818,991
; PRIOR FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 76
; SOFTWARE: Patenlin version 3.0
; SEQ ID NO 70
; LENGTH: 3458
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-277-161-70

```

Query Match	15.5%	Score 29.8	DB 12	Length 3458
Best Local Similarity	50.3%	Pred. No. 33		
Matches	73	Conservative 0	Mismatches 72	Indels 0
Gaps	0			
Qy	3	TCGCGCTACTGGTCTTTTCATATTTTGCATAGTCTTGTGGTGGCGCATCTTCCCGACA	62	
Db	1834	TCAGGTTCTTGTTCCTCATTTGGGAACATGGCTTTGTGACGGATAATCTTGAAACCCCG	1893	
Qy	63	TCATCTAAATTTGCTCTTTATTTGGTTTTTACGCCACATCTTGGCGATAAACAATATTCCGC	122	
Db	1894	TITCTAAATATTCTTTGGTCATTTAAGCCAACGGGAGGGTGGAAAAAGCCCTT	1953	
Qy	123	CTTGGCGTCGCGAATGTTCAAGCTA	147	
Db	1954	TATTTGCTCAGGAAGGTGATGTTA	1978	

```

RESULT 13
US-10-027-632-260325
, Sequence 260325, Application US/10027632
, Publication No. US20030204075A9
, GENERAL INFORMATION:
, APPLICANT: Wang, David G.
, TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
, TITLE OF INVENTION: Polymorphisms in the Human Genome
, FILE REFERENCE: 10827.129
, CURRENT APPLICATION NUMBER: US/10/027,632
, CURRENT FILING DATE: 2002-04-30
, PRIOR APPLICATION NUMBER: US 60/218,006
, PRIOR FILING DATE: 2000-07-12
, PRIOR APPLICATION NUMBER: US 60/198,676
, PRIOR FILING DATE: 2000-04-20
, PRIOR APPLICATION NUMBER: US 60/193,483
, PRIOR FILING DATE: 2000-03-29
, PRIOR APPLICATION NUMBER: US 60/185,218
, PRIOR FILING DATE: 2000-02-24
, PRIOR APPLICATION NUMBER: US 60/157,363
, PRIOR FILING DATE: 1999-11-23
, PRIOR APPLICATION NUMBER: US 60/156,358
, PRIOR FILING DATE: 1999-09-28
, PRIOR APPLICATION NUMBER: US 60/146,002
, PRIOR FILING DATE: 1999-08-09
, NUMBER OF SEQ ID NOS: 325720
, SOFTWARE: FastSeq for Windows Version 4.0
, SEQ ID NO 260325
, LENGTH: 3782
, TYPE: DNA
, ORGANISM: Human
US-10-027-632-260325

```

Query Match	15.4%;	Score 29.6;	DB 12;	Length 3782;
Best Local Similarity	50.7%;	Pred. No. 40;		
Matches	71;	Conservative 0;	Mismatches 69;	Indels 0; Gaps 0;
QY	11	CTTGGTTTTTCATATTTGCATAGTCTTGTCCGTGGGCACTTTCCTCCGACATCATCTAA	70	
Db	2273	CTTCCTTAATCATCTTCTCCCTCCCTGCCACCTCAGTCACCTCTGTGTCCTTTATCTGTG	2332	
QY	71	ATTTCGTCCTTATTGGCTTTTATGCCCACTCATTTCCGGATAAACAATATTCGGCTTGCCTG	130	
Db	2333	TTTTTGTTTTATCAGATATTCACCTCTTTTTCGGTGAAATATTATTATTATTTGCTGTT	2392	
QY	131	CGCGAATGTTCAAGCTAGCC	150	
Db	2393	TTCTAAATGTCAGTCTTGCC	2412	

RESULT 14  
US-10-027-632-260326  
; Sequence 260326, Application US/10027632  
; Publication No. US20030204075A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.



```
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 260326
; LENGTH: 3782
; TYPE: DNA
; ORGANISM: Human
; US-10-027-632-260326

Query Match          15.4%; Score 29.6; DB 12; Length 3782;
Best Local Similarity 50.7%; Pred. No. 40;
Matches 71; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 11 CTGTGTTTTTCATATTTTCATAGTCTTGTGGTGGGCATCTTCCCGACATCACTAA 70
Db 2273 CTTCCTTAATCATCTCTCCCTCCCTGCCACCTCAGTCACCCCTGTGTCCTTTATCTG 2332

QY 71 ATTTGCTTTTATGTTTTACGCCACTCATTTGGGATAAACAATATTCGCCCTTGCCGT 130
Db 2333 TTTTGTGTTTTATCACAGTATTCACCTCTTTTGGCGTGAATATTATTATTGTCGTG 2392

QY 131 CGCGAATGTTCAAGCTAGCC 150
Db 2393 TTCTAATTGTCAGTCTTGCC 2412

RESULT 15
US-10-027-632-260325
; Sequence 260325, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 260325
; LENGTH: 3782
; TYPE: DNA
```

```
; ORGANISM: Human
; US-10-027-632-260325

Query Match          15.4%; Score 29.6; DB 13; Length 3782;
Best Local Similarity 50.7%; Pred. No. 40;
Matches 71; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 11 CTGTGTTTTTCATATTTTCATAGTCTTGTGGTGGGCATCTTCCCGACATCACTAA 70
Db 2273 CTTCCTTAATCATCTCTCCCTCCCTGCCACCTCAGTCACCCCTGTGTCCTTTATCTG 2332

QY 71 ATTTGCTTTTATGTTTTACGCCACTCATTTGGGATAAACAATATTCGCCCTTGCCGT 130
Db 2333 TTTTGTGTTTTATCACAGTATTCACCTCTTTTGGCGTGAATATTATTATTGTCGTG 2392

QY 131 CGCGAATGTTCAAGCTAGCC 150
Db 2393 TTCTAATTGTCAGTCTTGCC 2412

Search completed: November 15, 2003, 08:31:48
Job time : 110.826 secs.
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 14, 2003, 23:56:24 ; Search time 531.201 seconds  
(without alignments)  
9241.609 Million cell updates/sec

Title: US-09-928-457-74  
Perfect score: 120  
Sequence: 1 CGGTGAGAACAGGCAAGCT.....CTCAAGGAGAGAGCTATCAT 120

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl.\*  
1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.cm.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.sts.\*  
12: gb.sy.\*  
13: gb.un.\*  
14: gb.vi.\*  
15: em.ba.\*  
16: em.fun.\*  
17: em.hum.\*  
18: em.in.\*  
19: em.mu.\*  
20: em.om.\*  
21: em.or.\*  
22: em.ov.\*  
23: em.pat.\*  
24: em.ph.\*  
25: em.pl.\*  
26: em.ro.\*  
27: em.sts.\*  
28: em.un.\*  
29: em.vi.\*  
30: em.htg.hum.\*  
31: em.htg.inv.\*  
32: em.htg.other.\*  
33: em.htg.mus.\*  
34: em.htg.pln.\*  
35: em.htg.rod.\*  
36: em.htg.mam.\*  
37: em.htg.vit.\*  
38: em.sy.\*  
39: em.hgo.hum.\*  
40: em.hgo.mus.\*  
41: em.hgo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	120	100.0	120	6	A68902	A68902 Sequence 73
2	120	100.0	120	6	A68903	A68903 Sequence 74
3	120	100.0	120	6	B063002	B063002 DNA and s
4	120	100.0	120	6	B063003	B063003 DNA and s
5	55.8	46.5	354	1	AF169447	AF169447 Neisseria
6	55.8	46.5	2094	6	AX024108	AX024108 Sequence
7	55.8	46.5	340806	1	NNA122491	AX024133 Sequence
8	55	45.8	2094	6	AX024133	AX024133 Sequence
9	55	45.8	10467	1	AE002381	AE002381 Neisseria
10	55	45.8	349980	6	AX043922	AX043922 Sequence
11	32.8	27.3	223203	2	AC113844	AC113844 Rattus no
12	31.8	26.5	174633	2	EX537269	EX537269 Dario rer
13	30.2	25.2	187986	2	AC138366	AC138366 Mus muscu
14	30.2	25.2	254098	2	AC096226	AC096226 Rattus no
15	30	25.0	204173	2	AC137237	AC137237 Rattus no
16	30	25.0	207918	2	AC112051	AC112051 Rattus no
17	30	25.0	258286	2	AC094841	AC094841 Rattus no
18	29.8	24.8	227385	2	AC094407	AC094407 Rattus no
19	29.8	24.8	235915	2	AC111641	AC111641 Rattus no
20	29.8	24.8	238377	2	AC125923	AC125923 Rattus no
21	29.8	24.8	240497	2	EX072570	EX072570 Dario rer
22	29.8	24.8	251180	2	AC123665	AC123665 Rattus no
23	29.8	24.8	279615	2	AC103332	AC103332 Rattus no
24	29.6	24.7	2635	8	AY098366	AY098366 Arabidops
25	29.6	24.7	2987	8	AY072210	AY072210 Arabidops
26	29.6	24.7	12519	1	AE006089	AE006089 Pasteurel
27	29.6	24.7	99254	8	AC002423	AC002423 Genomic s
28	29.6	24.7	199001	2	AC136115	AC136115 Rattus no
29	29.6	24.7	209574	2	AC102170	AC102170 Mus muscu
30	29.4	24.5	232018	2	AC125975	AC125975 Rattus no
31	29.4	24.5	230956	2	AC105782	AC105782 Rattus no
32	29.4	24.5	236290	2	AC114863	AC114863 Rattus no
33	29.4	24.5	239575	2	AC099194	AC099194 Rattus no
34	29.4	24.5	274620	2	AC126199	AC126199 Rattus no
35	29.2	24.3	10310	1	AE008086	AE008086 Agrobacte
36	29.2	24.3	10947	1	AE009121	AE009121 Agrobacte
37	29.2	24.3	140136	9	AL136087	AL136087 Human DNA
38	29.2	24.3	142711	10	AC121863	AC121863 Mus muscu
39	29.2	24.3	179668	10	AC124399	AC124399 Mus muscu
40	29.2	24.3	187191	9	AC007222	AC007222 Homo sapi
41	29.2	24.3	196181	2	AC144665	AC144665 Rattus no
42	29.2	24.3	221145	10	AC098739	AC098739 Mus muscu
43	29.2	24.3	237569	2	AC130141	AC130141 Rattus no
44	29.2	24.3	235653	2	AC110482	AC110482 Rattus no
45	29.2	24.3	237985	2	AC125754	AC125754 Rattus no

ALIGNMENTS

RESULT 1  
A68902 A68902 Sequence 73 from Patent WO9802547.  
LOCUS A68902  
DEFINITION A68902  
ACCESSION A68902  
VERSION A68902.1 GI:4759821  
KEYWORDS  
SOURCE unidentified  
ORGANISM unclassified  
REFERENCE 1 (bases 1 to 120)  
AUTHORS Nassif, X., Tinsley, C., Achtman, M., Ruelle, J., Vinals, C. and  
Meyer, P.  
TITLE DNA AND SPECIFIC PROTEINS OR PEPTIDES OF THE NEISSERIA MENINGITIDIS  
SPECIES BACTERIA, METHOD FOR OBTAINING THEM AND THEIR BIOLOGICAL

120 bp DNA linear PAT 06-MAY-1999

APPLICATIONS  
 Patent: WO 9802547-A 73 22-JAN-1998;  
 INST NAT SANTE RECH MED (FR)  
 Other Publication FR 2751000 19980116.  
 Location/Qualifiers  
 1. 120  
 /organism="unidentified"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:32644" 22 t  
 38 a 30 c 30 g

JOURNAL  
 source  
 BASE COUNT  
 ORIGIN  
 Query Match 100.0%; Score 120; DB 6; Length 120;  
 Best Local Similarity 100.0%; Pred. No. 6.3e-31;  
 Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CGGTGAGAACAGGCAAGTAAATGAAATGCTGAGGACGAGCTGTCTCGGAACGAAA 60  
 Db 1 CGGTGAGAACAGGCAAGTAAATGAAATGCTGAGGACGAGCTGTCTCGGAACGAAA 60  
 QY 61 ACTCCTTACCGAAGTCTTCTATACCCAGGCTCAATAGCGCTCAAGGAGAGGCTATCAT 120  
 Db 61 ACTCCTTACCGAAGTCTTCTATACCCAGGCTCAATAGCGCTCAAGGAGAGGCTATCAT 120

RESULT 2  
 A68903  
 LOCUS  
 DEFINITION  
 Sequence 74 from Patent WO9802547.  
 A68903  
 ACCESSION  
 VERSION  
 A68903.1 GI:4759822  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 unidentified  
 unclassified  
 unclassified.  
 Reference  
 Authors  
 Nassif,X., Tinsley,C., Achtman,M., Ruelle,J., Vinals,C. and  
 Merker,P.  
 Title  
 DNA AND SPECIFIC PROTEINS OR PEPTIDES OF THE NEISSERIA MENINGITIDIS  
 SPECIES BACTERIA, METHOD FOR OBTAINING THEM AND THEIR BIOLOGICAL  
 APPLICATIONS  
 Journal  
 Patent: WO 9802547-A 74 22-JAN-1998;  
 INST NAT SANTE RECH MED (FR)  
 Other Publication FR 2751000 19980116.  
 Location/Qualifiers  
 1. 120  
 /organism="unidentified"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:32644" 22 t  
 38 a 30 c 30 g 22 t

BASE COUNT  
 ORIGIN  
 Query Match 100.0%; Score 120; DB 6; Length 120;  
 Best Local Similarity 100.0%; Pred. No. 6.3e-31;  
 Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CGGTGAGAACAGGCAAGTAAATGAAATGCTGAGGACGAGCTGTCTCGGAACGAAA 60  
 Db 1 CGGTGAGAACAGGCAAGTAAATGAAATGCTGAGGACGAGCTGTCTCGGAACGAAA 60  
 QY 61 ACTCCTTACCGAAGTCTTCTATACCCAGGCTCAATAGCGCTCAAGGAGAGGCTATCAT 120  
 Db 61 ACTCCTTACCGAAGTCTTCTATACCCAGGCTCAATAGCGCTCAAGGAGAGGCTATCAT 120

RESULT 3  
 BD063002  
 LOCUS  
 DEFINITION  
 DNA and specific proteins or peptides of the Neisseria meningitidis  
 species bacteria, method for obtaining them and their biological  
 applications.  
 BD063002  
 ACCESSION  
 VERSION  
 BD063002.1 GI:22608505

KEYWORDS  
 SOURCE  
 ORGANISM  
 unclassified  
 unclassified  
 unclassified.  
 REFERENCE  
 1 (bases 1 to 120)  
 AUTHORS  
 Nassif,X., Tinsley,C., Achtman,M., Ruelle,J.L., Vinals,C. and  
 Merker,P.  
 TITLE  
 DNA and specific proteins or peptides of the Neisseria meningitidis  
 species bacteria, method for obtaining them and their biological  
 applications  
 JOURNAL  
 Patent: JP 2001504684-A 64 10-APR-2001;  
 INSTITUT NATIONAL DE LA SANTE ET DE LA RECHERCHE MEDICALE, MAX  
 PLANCK GESELLSCHAFT ZUR FORDERUNG DER WISSENSCHAFTEN EV BERLIN,  
 SMITHKLINE BEECHAM  
 COMMENT  
 PN JP 2001504684-A/64  
 PD 10-APR-2001  
 PF 11-JUL-1997 JP 1998505685  
 PR 12-JUL-1996 FR 96/08768  
 PI XAVIER NASSIF, COLIN TINSLEY, MARK ACHTMAN, JEAN LOUIS RUELLE, PI  
 CARLA VINALS.  
 PI PETRA MERKER  
 PC C12N15/31.C07K14/22.C07K16/12.A61K39/095.C12Q1/68.G01N33/53 CC  
 CC Strandedness: Single;  
 CC Topology: Linear;  
 FH Key Location/Qualifiers.  
 FEATURES  
 source  
 1. 120  
 /organism="unidentified"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:32644" 22 t  
 38 a 30 c 30 g 22 t  
 BASE COUNT  
 ORIGIN  
 Query Match 100.0%; Score 120; DB 6; Length 120;  
 Best Local Similarity 100.0%; Pred. No. 6.3e-31;  
 Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CGGTGAGAACAGGCAAGTAAATGAAATGCTGAGGACGAGCTGTCTCGGAACGAAA 60  
 Db 1 CGGTGAGAACAGGCAAGTAAATGAAATGCTGAGGACGAGCTGTCTCGGAACGAAA 60  
 QY 61 ACTCCTTACCGAAGTCTTCTATACCCAGGCTCAATAGCGCTCAAGGAGAGGCTATCAT 120  
 Db 61 ACTCCTTACCGAAGTCTTCTATACCCAGGCTCAATAGCGCTCAAGGAGAGGCTATCAT 120

RESULT 4  
 BD063003  
 LOCUS  
 DEFINITION  
 DNA and specific proteins or peptides of the Neisseria meningitidis  
 species bacteria, method for obtaining them and their biological  
 applications.  
 BD063003  
 ACCESSION  
 VERSION  
 BD063003.1 GI:22608606  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 unclassified  
 unclassified  
 unclassified.  
 REFERENCE  
 1 (bases 1 to 120)  
 AUTHORS  
 Nassif,X., Tinsley,C., Achtman,M., Ruelle,J.L., Vinals,C. and  
 Merker,P.  
 TITLE  
 DNA and specific proteins or peptides of the Neisseria meningitidis  
 species bacteria, method for obtaining them and their biological  
 applications  
 JOURNAL  
 Patent: JP 2001504684-A 65 10-APR-2001;  
 INSTITUT NATIONAL DE LA SANTE ET DE LA RECHERCHE MEDICALE, MAX  
 PLANCK GESELLSCHAFT ZUR FORDERUNG DER WISSENSCHAFTEN EV BERLIN,  
 SMITHKLINE BEECHAM  
 COMMENT  
 PN JP 2001504684-A/65  
 PD 10-APR-2001  
 PF 11-JUL-1997 JP 1998505685  
 PR 12-JUL-1996 FR 96/08768  
 PI XAVIER NASSIF, COLIN TINSLEY, MARK ACHTMAN, JEAN LOUIS RUELLE, PI

```

CARLA VINALS,
PI PETRA MERKER
PC C12N15/31,C07K14/22,C07K16/12,A61K39/095,C12Q1/68,G01N33/53 CC
Strandedness: Single;
CC Topology: Linear;
FH Key Location/Qualifiers.

FEATURES
    source
        1..120
            /organism="unidentified"
            /mol_type="genomic DNA"
            /db_xref="taxon:32644"
BASE COUNT      38 a      30 g      22 t
ORIGIN
Query Match      100.0%; Score 120; DB 6; Length 120;
Best Local Similarity 100.0%; Pred. No: 6.3e-31;
Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGTGAGAACGAGCAAGTAAATGAAATGCTTGAGGACGAGCTGTGTCGGAACGAAA 60
Db 1 CGGTGAGAACGAGCAAGTAAATGAAATGCTTGAGGACGAGCTGTGTCGGAACGAAA 60

QY 61 ACTCCTTACCGAAGTCTTCTATACCCAGGCTCAATAGCGGCTCAAGGAGAGCTATCAT 120
Db 61 ACTCCTTACCGAAGTCTTCTATACCCAGGCTCAATAGCGGCTCAAGGAGAGCTATCAT 120

RESULT 5
AF169447
LOCUS      AF169447      354 bp      DNA      linear      BCT 09-AUG-2000
DEFINITION      Neisseria meningitidis strain Z2491 clone Cm020 unknown sequence.
ACCESSION      AF169447
VERSION      AF169447.1 GI:9754655
KEYWORDS
SOURCE      Neisseria meningitidis
ORGANISM      Neisseria meningitidis
Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
Neisseriaceae; Neisseria.
REFERENCE      1 (bases 1 to 354)
AUTHORS      Perrin,A., Nassif,X. and Tinsley,C.R.
TITLE      Identification of regions of the chromosome of Neisseria
            meningitidis and Neisseria gonorrhoeae which are specific to
            pathogenic Neisseriae
JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 354)
AUTHORS      Perrin,A., Nassif,X. and Tinsley,C.R.
TITLE      Direct Submission
JOURNAL      Submitted (16-JUL-1999) Necker Medicine Faculty, INSERM U411, 156
            rue de Vaugirard, Paris 75015, France
            Location/Qualifiers
                1..354
                    /organism="Neisseria meningitidis"
                    /mol_type="genomic DNA"
                    /strain="Z2491"
                    /db_xref="taxon:487"
                    /clone="Cm020"
BASE COUNT      122 a      77 c      81 g      73 t      1 others
ORIGIN
Query Match      46.5%; Score 55.8; DB 1; Length 354;
Best Local Similarity 79.8%; Pred. No: 2.2e-08;
Matches 79; Conservative 0; Mismatches 17; Indels 3; Gaps 1;

QY 17 AGTAATGAAATGCTGAGGACGAGCTGTGCTCGAACGAAAACTCTTACCG---AA 73
Db 20 AGTAATGAAATGCTGAGGACGAGCTGTGCTCGAACGAAAACTCTTACCCACAA 79

QY 74 GTCTTCTATACCCAGGCTCAATAGCGGCTCAAGGAGAGA 112
Db 80 GACTTCTATACCCAGGCTCAATAGCGGCTCAAGGAGAGA 118

RESULT 6
AF169447
LOCUS      AF169447      2094 bp      DNA      linear      PAT 15-SEP-2000
DEFINITION      Sequence 51 from Patent FR2785293.
ACCESSION      AX024108
VERSION      AX024108.1 GI:10184420
KEYWORDS      Neisseria meningitidis
SOURCE      Neisseria meningitidis
ORGANISM      Neisseria meningitidis
Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
Neisseriaceae; Neisseria.
REFERENCE      1
AUTHORS      Nassif,X., Tinsley,C., Aujame,L., Perrin,A., Rokbi,B.,
            Bouchardon,A. and Renaud,M.C.
JOURNAL      Patent: FR 2785293-A 51 05-MAY-2000;
            PASTEUR MERIEUX SERUMS VACC (FR)
            Location/Qualifiers
                1..2094
                    /organism="Neisseria meningitidis"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:487"
                1..2094
                    /note="unnamed protein product"
                    /codon_start=1
                    /transl_table=1
                    /protein_id="CAC08930.1"
                    /db_xref="GI:10184421"
                    /translation="MNSTASKTLKGLSLVFFASGFCALIQVSWQRLLESFHSIGLIDLS
            ITVIISFVMVGLGVGAYFGRIADRPSSIIPLEFCIAEVSIGLGLVSRGLSGLGHL
            LVEADLPPIIAANFILLPLTFMNGATLPLTFCFNKIHNVGESIGTLTYFFNTLGA
            LGLAAAEFFVFFTLQITALTACFNLLIAASVCCVTERMDIVNTKPTSLYMLSP
            LSGLLGLEVLWVRMFSPOAQVPOAFTFLAYFLTGIAGVAYFGKRCRSFVDIP
            FIGCFLWAGIADFLILGAALLTSGSFVHAGIFITLSAVVRGLIFPLVHVGTG
            NKSGRVSNVYFANVAGSALGPVLIGFVLDFLSTQIILLICLIANAAPLPTLFOK
            SLRNVAVSVMFGLMFLPDPSVFQNIADRPDLIENKGIIVAVHRDGGKVYIG
            ANVDGAYNTDVFNSVNGIERAYLLPSLKGIRIFVWGLSTGSRWARVLSAIPQMOSM
            IVAENFAYSLADPEQAPLLQDXRVEIVLDGKRLRRHDPEDFDLIMNTTWY
            RAYSTNLLSAEFLKQVSHLTDPGIVMFNTHSPHAFATAVHSIPVAYRYGHMVGSA
            TPVFFNKKELLKORSLWPESGRHVFDSSTVDAAAQKVSRMLIQMTEPSAGAEVI
            TDDNMLVEYKVRGI"
BASE COUNT      428 a      500 c      512 g      654 t
ORIGIN
Query Match      46.5%; Score 55.8; DB 6; Length 2094;
Best Local Similarity 79.8%; Pred. No: 2e-08;
Matches 79; Conservative 0; Mismatches 17; Indels 3; Gaps 1;

QY 17 AGTAATGAAATGCTGAGGACGAGCTGTGCTCGAACGAAAACTCTTACCG---AA 73
Db 747 AGTAATGAAATGCTGAGGACGAGCTGTGCTCGAACGAAAACTCTTACCCACAA 688

QY 74 GTCTTCTATACCCAGGCTCAATAGCGGCTCAAGGAGAGA 112
Db 687 GACTTCTATACCCAGGCTCAATAGCGGCTCAAGGAGAGA 649

RESULT 7
NMA122491
LOCUS      NMA122491      340806 bp      DNA      linear      BCT 02-SEP-2002
DEFINITION      Neisseria meningitidis serogroup A strain Z2491 complete genome;
            segment 1/7
VERSION      AL162752.2 GI:7378778
KEYWORDS      Neisseria meningitidis Z2491
SOURCE      Neisseria meningitidis Z2491
ORGANISM      Neisseria meningitidis Z2491
Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
Neisseriaceae; Neisseria.
REFERENCE      1 (bases 1 to 340806)
AUTHORS      Parkhill,J., Achtman,M., James,K.D., Bentley,S.D., Churcher,C.,
            Klee,S.R., Morelli,G., Basham,D., Brown,D., Chillingworth,T.,
            Davies,R.M., Davis,P., Devlin,K., Feltwell,T., Hamlin,N.,
            Holroyd,S., Jagels,K., Leather,S., Moule,S., Mungall,K.,
            Quail,M.A., Rajandream,M.A., Rutherford,K.M., Simmonds,M.,

```

Skelton, J., Whitehead, S., Spratt, B.G. and Barrell, B.G.  
Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491

Nature 404 (6777), 502-506 (2000)

20222556  
MEDLINE  
PUBMED  
10761919

REFERENCE  
2 (bases 1 to 340806)

Parkhill, J.

Direct Submission

Submitted (30-Mar-2000) Submitted on behalf of the Neisseria sequencing team, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk

Notes:

Details of *N. meningitidis* sequencing at the Sanger Centre are available on the World Wide Web.

(URL: [http://www.sanger.ac.uk/Projects/N\\_meningitidis/](http://www.sanger.ac.uk/Projects/N_meningitidis/)).

Location/Qualifiers

1..340806  
/organism="Neisseria meningitidis Z2491"

/mol\_type="genomic DNA"

/strain="Z2491"

/db\_xref="taxon:122587"

/note="serogroup: A"

complement(38..358)

/gene="NMA0001"

complement(38..358)

/gene="NMA0001"

/note="NMA0001, unknown, len: 106 aa"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein NMA0001"

/protein\_id="CAB83321.1"

/db\_xref="GI:7378779"

/db\_xref="SPTREMBL:Q9JX93"

/translation="MSNTQTIRSDFTTSRAWGLDIAWMNGTTRVLRHWTDQPKYKHV  
NDGEVFVMDGCEVDMHYRENGEHEIVLKSGLDIFAGIGCTEHVHPRGRTRILVIEK  
EGSV"

complement(414..2438)

/gene="nuoL"

complement(414..2438)

/gene="nuoL"

/EC\_number="1.6.5.3"

/note="NMA0002, nuoL, NADH dehydrogenase I chain L, len: 674 aa; similar to many e.g. NUOL\_RHOCA NADH dehydrogenase I chain L (EC 1.6.5.3) (700 aa). Contains Pfam match to entry PF00361 oxidored q1, NADH-ubiquinone/plastoquinone (complex I), various chains, and to entry PF00662 oxidored q1\_N, NADH-ubiquinone oxidoreductase (complex I), chain 5 N-terminus"

/codon\_start=1

/transl\_table=11

/product="NADH dehydrogenase I chain L"

/protein\_id="CAB83322.1"

/db\_xref="GI:7378780"

/db\_xref="SWISS-PROT:Q9JX92"

/translation="MNDTLYLIILALVPLAGSLIAGFGNKIGRAGAHHTVILGVAVS  
AVLSAYLWGLNGSRKAFDENVTYLTMTGGLDFSGVLTDTMTAMMMVVVTGVSVMV  
HYTGYVHDKRGVQRFYSYISLFTSMLMLINSNNFIQLFGWEAVGLVSYLLIGF  
VFKRPSATFANLIRNVGDGFLGIGVLAVYFGSLRYQDVPAVLNVQNTATIQ  
LFGVEMSLITVCTLLPVGAMGSAQFPLHVLDPDSMEGPTPLSALIHAAVTYAGL  
FWSRMSIYEMSSTALSIVNIGAITALEMGLFCIQNDIKRVAYSTLSQLGYMTV  
ALGASAYVMVPHVHTAFKALLFLAAGSAIIGMHHDQDMHMGNLKKYMPITWLTM  
LIGNSLGTFFPGFYKSDSIEAAKYSTLPGSFPAYFAVLASVFTAFYARQYFM  
VFGHEKRSPLPEHSDGHEHGLKNDNPHESPLVTLPLILLAVPSVVIIGYIAI  
EPMYGDSPFDKQVIFVNAADHTMHMKEEFHGALAMVSHSLSPVLALAGVLSAWL  
LYVKLPHLPKIAQAFREPVYVLFENKYVLDALYENFAKGTALGTFFKVGDTAID  
NGVNGSARLNCVAAQVRKVQTGFIYTYAAAMVFGVILLGMTFFWGLER"

1035..1044

/note="Core DNA uptake sequence: gccgtctgaa"

/label=DUS

complement(1116..2030)

/gene="nuoL"

/note="Pfam match to entry PF00361 oxidored q1,

misc\_feature

/gene="nuoL"

/note="Pfam match to entry PF00662 oxidored q1\_N,

NADH-ubiquinone oxidoreductase (complex I), chain 5

N-terminus, score 77.20, E-value 3.5e-19"

complement(2487..2828)

/gene="NMA0003"

complement(2487..2828)

/gene="NMA0003"

/note="NMA0003, unknown, len: 113 aa"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein NMA0003"

/protein\_id="CAB83323.1"

/db\_xref="GI:7378781"

/db\_xref="SPTREMBL:Q9JX91"

/translation="NMKNQLEQFHKAMLIYQALNLPQYKATRFLOIVNERGGK  
EADKLLSTGKTKTOTGTLLSGGVHALKYSMEYLVLOKPKWCDLFTFEQLAVARK  
RLRVGFVSPK"

complement(2833..2837)

/gene="NMA0004"

complement(2952..3527)

/gene="NMA0004"

complement(2952..3527)

/gene="NMA0004"

/note="NMA0004, unknown, len: 191 aa; similar to hypothetical proteins e.g. Y977\_HAEIN HI9777 (191 aa), fasta scores; E(): 0; 80.5% identity in 190 aa overlap, and to proteins involved in cell cycle e.g. TR:O54679 (EMBL:AF036487) Lactococcus lactis Plasmid pNZ4000

putative mobilization protein (200 aa), fasta scores; E(): 0.55.1% identity in 167 aa overlap, and FIC ECOLI cell filamentation protein FIC (200 aa), fasta scores; E(): 0.093, 29.4% identity in 126 aa overlap"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein NMA0004"

/protein\_id="CAB83324.1"

/db\_xref="GI:7378782"

/db\_xref="SPTREMBL:Q9JQR9"

/translation="MPSENPIKTKWKSIDQSLHNNARLFPESGDDIRIEVSTAGLQQ  
IHRYLFGGLYFAGQIDENISKGFNFANAYLKEALVKIEQMPERFETIAKYVE  
NNIAHPFLGNGRSTRIDLVKKNLKVNNQVSKTYLQAMERSFVNDLELRFLL  
LADNLTDDVDNREIIFKQIEQSYYYEYEGK"

complement(3516..3525)

/gene="NMA0004"

/note="Core DNA uptake sequence: gccgtctgaa"

/label=DUS

complement(3561..3866)

/gene="nuoK"

complement(3561..3866)

/gene="nuoK"

/EC\_number="1.6.5.3"

/note="NMA0005, nuoK, NADH dehydrogenase I chain K, len: 101 aa; similar to many e.g. NUOK\_RHOCA NADH dehydrogenase I chain K (EC 1.6.5.3) (102 aa), fasta scores; E(): 1.1e-24; 65.0% identity in 100 aa overlap. Contains Pfam match to entry PF00420 oxidored q2,

NADH-ubiquinone/plastoquinone oxidoreductase chain 4L"

/codon\_start=1

/transl\_table=11

/product="NADH dehydrogenase I chain K"

/protein\_id="CAB83325.1"

/db\_xref="GI:7378783"

/db\_xref="SPTREMBL:Q9JQU9"

/translation="MITLHYLVLCALLFGISAMGIFMNRKNVLVLLMSIELMLLVAN  
FNFIASQHLGDTAQIIFVFVLTVAAESALGLAIMLVYVNRQTINNVADLDELKG"

complement(3564..3863)

/gene="nuoK"

/note="Pfam match to entry PF00420 oxidored q2,

NADH-ubiquinone/plastoquinone oxidoreductase chain 4L,

score 141.50, E-value 1.5e-38"

misc\_feature

misc\_feature

```

Gene      complement (3863. .4534)
CDS       /gene="nuoJ"
          /complement (3863. .4534)
          /gene="nuoJ"
          /EC_number="1.6.5.3"
          /note="NMA0006, nuoJ, NADH dehydrogenase I chain J, len:
          223 aa; similar to many e.g. NUOJ, RHOCA NADH dehydrogenase
          I chain J (EC 1.6.5.3) (202 aa), Fatsa scores: E():
          9.7e-21; 41.3% identity in 201 aa overlap. Contains Pfam
          match to entry PF00499 oxidored q3,
          NADH-ubiquinone/plastoquinone oxidoreductase chain 6"
          /codon_start=1
          /transl_table=11
          /product="NADH dehydrogenase I chain J"
          /protein_id="CAB83326.1"
          /db_xref="GI:7378784"
          /db_xref="SPTREMBL:O9JX90"
          /translation="MTFQILFYIFAVILLYGAIKTVAKNPVHAALHLVTFVCVSAM
          LMLMQAFLGVTLVVVGVAVNVLFFVVMMLNIDIEEMRAGFWFRAHPAVAGVVGTL
          AVALILIVNPKTDAAFGMLMKDIPADYNNIRDLGSRTYTDLYLFFELAAVLLGMV
          RAIALVHRKTVNPKMDPADQVKVRADQGRMLVKMEAVKPTESAEESEVSDDLKPK
          BEGKA"
          /gene="nuoJ"
          /complement (3874. .3879)
          /gene="nuoJ"
          /complement (4034. .4525)
          /gene="nuoJ"
          /note="Pfam match to entry PF00499 oxidored q3,
          NADH-ubiquinone/plastoquinone oxidoreductase chain 6,
          score 77.90, E-value 2e-19"
          /complement (4570. .5277)
          /gene="NMA0007"
          /complement (4570. .5277)
          /gene="NMA0007"
          /note="NMA0007, unknown, len: 235 aa"
          /codon_start=1
          /transl_table=11
          /product="hypothetical protein NMA0007"
          /protein_id="CAB83327.1"

Query Match      46.5%; Score 55.8; DB 1; Length 340806;
Best Local Similarity 79.8%; Pred. No. 1.5e-08;
Matches 79; Conservative 0; Mismatches 17; Indels 3; Gaps 1;

QY 17 AGTATGAATGCTGAGGACGACTGTCTCGACGAAACTCTTACCG---AA 73
Db 17124 ATTAATGAATGCTGAGGACGACTGTCTCGACGAAACTCTTACCGCAA 17183

QY 74 GTCTTATACCGAGCTCAATAGCGGCTCAAGGAGAGA 112
Db 17184 GACTTATACCAAGCTCAATAGCGGCTCAAGGAGAGA 17222

RESULT 8
AX024133/c
LOCUS      AX024133      2094 bp      DNA      linear      PAT 15-SEP-2000
DEFINITION      Sequence 76 from Patent FR2785293.
ACCESSION      AX024133
VERSION        AX024133.1
KEYWORDS
SOURCE
ORGANISM      Neisseria gonorrhoeae
              Neisseria gonorrhoeae
              Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
              Neisseriaceae; Neisseria.
REFERENCE
1
AUTHORS      Nassif, X., Tinsley, C., Aujane, L., Perrin, A., Rokbi, B.,
              Bouchardon, A. and Renaud, M.G.
JOURNAL      Patent: FR 2785293-A 76 05-MAY-2000;
              PASTEUR MERIEUX SERUMS VACC (FR)
FEATURES
            Location/Qualifiers
            source
              1. .2094
              /organism="Neisseria gonorrhoeae"
              /mol_type="genomic DNA"
              /db_xref="taxon:485"

complement (3863. .4534)
/gene="nuoJ"
/complement (3863. .4534)
/gene="nuoJ"
/EC_number="1.6.5.3"
/note="NMA0006, nuoJ, NADH dehydrogenase I chain J, len:
223 aa; similar to many e.g. NUOJ, RHOCA NADH dehydrogenase
I chain J (EC 1.6.5.3) (202 aa), Fatsa scores: E():
9.7e-21; 41.3% identity in 201 aa overlap. Contains Pfam
match to entry PF00499 oxidored q3,
NADH-ubiquinone/plastoquinone oxidoreductase chain 6"
/codon_start=1
/transl_table=11
/product="NADH dehydrogenase I chain J"
/protein_id="CAB83326.1"
/db_xref="GI:7378784"
/db_xref="SPTREMBL:O9JX90"
/translation="MTFQILFYIFAVILLYGAIKTVAKNPVHAALHLVTFVCVSAM
LMLMQAFLGVTLVVVGVAVNVLFFVVMMLNIDIEEMRAGFWFRAHPAVAGVVGTL
AVALILIVNPKTDAAFGMLMKDIPADYNNIRDLGSRTYTDLYLFFELAAVLLGMV
RAIALVHRKTVNPKMDPADQVKVRADQGRMLVKMEAVKPTESAEESEVSDDLKPK
BEGKA"
/gene="nuoJ"
/complement (3874. .3879)
/gene="nuoJ"
/complement (4034. .4525)
/gene="nuoJ"
/note="Pfam match to entry PF00499 oxidored q3,
NADH-ubiquinone/plastoquinone oxidoreductase chain 6,
score 77.90, E-value 2e-19"
/complement (4570. .5277)
/gene="NMA0007"
/complement (4570. .5277)
/gene="NMA0007"
/note="NMA0007, unknown, len: 235 aa"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA0007"
/protein_id="CAB83327.1"

Query Match      46.5%; Score 55.8; DB 1; Length 340806;
Best Local Similarity 79.8%; Pred. No. 1.5e-08;
Matches 79; Conservative 0; Mismatches 17; Indels 3; Gaps 1;

QY 17 AGTATGAATGCTGAGGACGACTGTCTCGACGAAACTCTTACCG---AA 73
Db 17124 ATTAATGAATGCTGAGGACGACTGTCTCGACGAAACTCTTACCGCAA 17183

QY 74 GTCTTATACCGAGCTCAATAGCGGCTCAAGGAGAGA 112
Db 17184 GACTTATACCAAGCTCAATAGCGGCTCAAGGAGAGA 17222

RESULT 8
AX024133/c
LOCUS      AX024133      2094 bp      DNA      linear      PAT 15-SEP-2000
DEFINITION      Sequence 76 from Patent FR2785293.
ACCESSION      AX024133
VERSION        AX024133.1
KEYWORDS
SOURCE
ORGANISM      Neisseria gonorrhoeae
              Neisseria gonorrhoeae
              Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
              Neisseriaceae; Neisseria.
REFERENCE
1
AUTHORS      Nassif, X., Tinsley, C., Aujane, L., Perrin, A., Rokbi, B.,
              Bouchardon, A. and Renaud, M.G.
JOURNAL      Patent: FR 2785293-A 76 05-MAY-2000;
              PASTEUR MERIEUX SERUMS VACC (FR)
FEATURES
            Location/Qualifiers
            source
              1. .2094
              /organism="Neisseria gonorrhoeae"
              /mol_type="genomic DNA"
              /db_xref="taxon:485"

1. .2094
/note="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAB08942.1"
/db_xref="GI:10184445"
/translation="MNSTASKTLKGLSVFFASGFCALIQVSWQRLLFSGHIGIDLS
ITVIVSVFVGLGVGAYFGRIADRPSSIIPLFCIAEVSIGLFGVSKGLISGLGHL
LVADLPITAAANFLLLPTFMWGATLLPTCFNKRKHNNVGESIGTLYFFNTLGA
LGSLAAEFVVFETTSOIIATACINLLIAASVCCVTERMDVNTKPTSVNMLSP
LTGSLGLIEVLWVRFSAQSVPGAFSIFLACFLTGAGVAFKKICRSRFDVDP
FTGCFWAGIADFLILGAALLTGFSGFHHAGIFITLSAVVRGIFLFEVHVGVDG
NKSGRQSVNYFANVAGSALGPVLIGVILDLSTQOIIYLLICLISAAVFLCTFQK
SLRNSVAVSVLMFGLIMFLPDSVFQNIAGRPDLRIENKHGIVAVYHDEGKVYV
ANYDGAINTDIENSNGIERAYLLPSLKGIRRIEFGVGLSTGSWAVLSAIPMQSM
IVAEINPAYRSIIADPEIAPLQDKRVEIIVDDGKRLRRHPDERFDLILMNSTWY
RAYSTNLLSGLFLKQVSHLTDPGIYVNTTHSPHAFATVHSIPYAVYGHMVGSA
TPVPFNKLLKQRLSLWPESGRHVFDSSIVDAAAKQVSRMLRMTEPESGAEVI
TDNNIVEVYKIGRI"
BASE COUNT      428 a 495 c 515 g 656 t
ORIGIN
Query Match      45.8%; Score 55; DB 6; Length 2094;
Best Local Similarity 81.1%; Pred. No. 3.8e-08;
Matches 77; Conservative 0; Mismatches 15; Indels 3; Gaps 1;

QY 21 AATGAAATCCCTGAGCAGGACTGTCTGCGAAGAAACTCTTACCG---AAGTCT 77
Db 743 AATGAAATCCCTGAGCAGGACTGTCTGCGAAGAAACTCTTACCGCAAGACT 684

QY 78 TCTATACCGAGCTCAATAGCGGCTCAAGGAGAGA 112
Db 683 TCTATACCGAGCTCAATATCCGGTAAGGAGAGA 649

RESULT 9
AE002381/c
LOCUS      AE002381      10467 bp      DNA      linear      BCT 25-MAY-2000
DEFINITION      Neisseria meningitidis serogroup B strain MC58 section 23 of 206 of
              the complete genome.
ACCESSION      AE002381
VERSION        AE002381.1
KEYWORDS      AE002381.1 GI:7225455
SOURCE
ORGANISM      Neisseria meningitidis MC58
              Neisseria meningitidis MC58
              Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
              Neisseriaceae; Neisseria.
REFERENCE
1 (bases 1 to 10467)
Tettelin, H., Saunders, N.J., Heidelberg, J., Jeffries, A.C.,
Nelson, K.E., Eisen, J.A., Ketchum, K.A., Hood, D.W., Peden, J.F.,
Dodson, R.J., Nelson, W.C., Gwinn, M.L., DeBoy, R., Peterson, J.D.,
Hickey, E.K., Haft, D.H., Salzberg, S.L., White, O., Fleischmann, R.D.,
Dougherty, B.A., Mason, T., Ciecko, A., Parksey, D.S., Blair, E.,
Cittone, H., Clark, B.B., Cotton, M.D., Utterback, T.R., Khouri, H.,
Qin, H., Vamathevan, J., Gill, J., Scarlato, V., Maignani, V.,
Pizza, M., Grandi, G., Sun, L., Smith, H.O., Fraser, C.M., Moxon, E.R.,
Rappuoli, R. and Venter, J.C.
Complete genome sequence of Neisseria meningitidis serogroup B
strain MC58
Science 287 (5459), 1809-1815 (2000)
JOURNAL
MEDLINE
PUBMED      20175755
REFERENCE
2 (bases 1 to 10467)
Tettelin, H., Saunders, N.J., Heidelberg, J., Jeffries, A.C.,
Nelson, K.E., Eisen, J.A., Ketchum, K.A., Hood, D.W., Peden, J.F.,
Dodson, R.J., Nelson, W.C., Gwinn, M.L., DeBoy, R., Peterson, J.D.,
Hickey, E.K., Haft, D.H., Salzberg, S.L., White, O., Fleischmann, R.D.,
Dougherty, B.A., Mason, T., Ciecko, A., Parksey, D.S., Blair, E.,
Cittone, H., Clark, B.B., Cotton, M.D., Utterback, T.R., Khouri, H.,
Qin, H., Vamathevan, J., Gill, J., Scarlato, V., Maignani, V.,
Pizza, M., Grandi, G., Sun, L., Smith, H.O., Fraser, C.M., Moxon, E.R.,
Rappuoli, R. and Venter, J.C.
Direct Submission
TITLE

```

JOURNAL Submitted (17-MAR-2000) The Institute for Genomic Research, 9712  
 Medical Center Dr, Rockville, MD 20850, USA  
 FEATURES  
 Location/Qualifiers  
 1. .10467  
 /organism="Neisseria meningitidis MC58"  
 /mol\_type="genomic DNA"  
 /strain="MC58"  
 /db\_xref="taxon:122586"  
 /note="serogroup: B"  
 69. .170  
 /gene="NMB0233"  
 CDS  
 69. .170  
 /gene="NMB0233"  
 /note="hypothetical protein; identified by Glimmer2;  
 putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="AAF40688.1"  
 /db\_xref="GI:7225456"  
 /translation="MGPFYESSPKQTAFLAVIIVDEAHATMYTQSK"  
 517. .1056  
 /gene="NMB0234"  
 CDS  
 517. .1056  
 /gene="NMB0234"  
 /note="hypothetical protein; identified by Glimmer2;  
 putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="AAF40689.1"  
 /db\_xref="GI:7225457"  
 /translation="MSQVFKDFDLSSVWKTNSWADENYKEAFTPEILAAVESLGYK  
 LPOSTIELMAVNGGIFVKNGCFPTTORSWAENHVOICEVSGIGFEKSGICGAMGOK  
 LWELEWYPPGIVGVPANDPSSGHAMFALDIYACDKQEPKPVFVEQSDPEIYELAPD  
 FEIFIRSRHDEFDIEEI"  
 complement(1297. .1458)  
 /gene="NMB0235"  
 CDS  
 complement(1297. .1458)  
 /gene="NMB0235"  
 /note="hypothetical protein; identified by Glimmer2;  
 putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="AAF40690.1"  
 /db\_xref="GI:7225458"  
 /translation="WNIINYKGGDDIISLTREELQLRLSLVIEIYAGVCIDABBEFIV  
 SGIRNPKLV"  
 complement(1466. .1669)  
 /gene="NMB0236"  
 CDS  
 complement(1466. .1669)  
 /gene="NMB0236"  
 /note="hypothetical protein; identified by Glimmer2;  
 putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="AAF40691.1"  
 /db\_xref="GI:7225459"  
 /translation="MLKRNLGFTDFKLITIFFLGIRLQSKHRAIWDDKSGAIVIRD  
 PMSKDGGAFTPLGKTYFDKQK"  
 complement(1582. .1941)  
 /gene="NMB0237"  
 CDS  
 complement(1582. .1941)  
 /gene="NMB0237"  
 /note="hypothetical protein; identified by Glimmer2;  
 putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="AAF40692.1"  
 /db\_xref="GI:7225460"

gene  
 1994. .2258  
 /gene="NMB0238"  
 /note="This region contains a gene with one or more  
 premature stops or frameshifts, and is not the result of a  
 sequencing artifact; similar to GB:I42023 PID:1007326  
 PID:1205567 SP:Q57255 PID:1574795 percent identity:  
 100.00; identified by sequence similarity; p"  
 2457. .3077  
 /gene="NMB0239"  
 CDS  
 2457. .3077  
 /gene="NMB0239"  
 /note="hypothetical protein; identified by Glimmer2;  
 putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="AAF40693.1"  
 /db\_xref="GI:7225461"  
 /translation="MNSTASKTLKGLSLVFFASGFCALYQVSWORLLPSHIGIDLSS  
 ITVIISVFMVIGVGAYFGRIADRPSSIIPLFCIAEVSIGLFGVIRGSLISGHL  
 LVEADLPITAAANFLLLPTFMGATLPLTCFNRKHNVGESIGLYLFNTLGA  
 LGSAAAEFFVFFTLSTQIALTACFNLLIAASVWLYRKDCGYSEH"  
 3100. .4551  
 /gene="NMB0240"  
 CDS  
 3100. .4551  
 /gene="NMB0240"  
 /note="hypothetical protein; identified by Glimmer2;  
 putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="AAF40694.1"  
 /db\_xref="GI:7225462"  
 /translation="MLSFLSGLLSGLIEVLVWRMSPFAAQVQAFSTLACFLTGLIA  
 VGAYGKRICRFRVDIPIGQCFWAGIADFLILGAALITFGSGFVHAGIFITLS  
 AVVRGLIFPLVHRHVDGNGKQSVNYFANVAGSALGPLVILDFLSTQIYL  
 LICLSAAYPLFTLFOKSLRLNVAVLMFGLMELPDSVPONLADRPDLIEN  
 KHGIYAVHRDGDKVYGVANVYDGAIVTDVNSVNGIERAYLLPSLKSIGIRIFVGL  
 STGSHARVLSAPEMQSMIVAEINPAYRSLIADDEQIAPLQDKRVEIVLDGKWL  
 RHPDEFDLINMTTWIRAYSLNLSAEFLQVQSHUTPDGIVNFTNTHFAPATA  
 VHSIPYAYRYGHMVGSATPVFPNKELKORLSLWPSGRHVFDSSTVDAQAQV  
 VSRMLIQTEPSAGAEVITDDNMIVEYKGRGI"  
 5013. .5369  
 /gene="NMB0241"  
 CDS  
 5013. .5369  
 /gene="NMB0241"  
 /note="similar to GP:3025801 percent identity: 80.51;  
 identified by sequence similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="NADH dehydrogenase I, A subunit"  
 /protein\_id="AAF40695.1"  
 /db\_xref="GI:7225463"  
 /translation="MLSAYFPVFFILIGLAGVLFILLGLQPKRHYAKDAPYEC  
 GFEAFENARKDFEYLYVAILFILFDLEAVMLPWAUVFKDLGAYGFWMLFIVVL  
 TVGFVYEWKKGALWE"  
 5360. .5842  
 /gene="NMB0242"  
 CDS  
 5360. .5842  
 /gene="NMB0242"  
 /note="similar to PID:2258365 percent identity: 90.06;  
 identified by sequence similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="NADH dehydrogenase I, B subunit"  
 /protein\_id="AAF40696.1"  
 /db\_xref="GI:7225464"  
 /translation="MGISGLVKKGPIITTSADTVLNYMPTGSLWPTVFLGACCAVEMMH  
 AGMAYDLORFGIIFRPSQROADMIVAGITNKNAPALRRVYDGLAEFRVLSMGSC  
 ANGGGYHYISVSVRGADRVVDVYVPCPTAEALTYGLIQLOOKIKRTSTIARDE



```

"
5855. 6448
/gene="NMB0243"
5855. 6448
/gene="NMB0243"
/note="similar to GP.3282561 percent identity: 69.41;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="NADH dehydrogenase I, C subunit"
/protein_id="RAFA0697.1"
/db_xref="GI:7225463"
/translation="MASIQDLYETVSRVLGNQAGKVISALGEIVTECLPHEYSVMTA
LRDEHDFELLPLDGVYDSTYKNEAQWGRFAVQSLLSKNQIRVYVWVSDDD
FPVSVSDIYNISADWVEREAFDYGIMFNHNPDLRLITDYGFGVHPFRKDFPISGY
VEMRYDEQKRVIPQVTIEPRITPRIVRENYGGQ"
6438. 7694
/gene="NMB0244"

Query Match 45.8%; Score 55; DB 1; Length 10467;
Best Local Similarity 81.1%; Pred. No. 3.5e-08;
Matches 77; Conservative 0; Mismatches 15; Indels 3; Gaps 1;

QY 21 AATGAATGCTGTAGGCACGACTGTCTGCGAACGAAACCTCTTACCG---AAGTCT 77
|||||
DB 3200 AATGAATGCTGTAGGCACGACTGTCTGCGAACGAAACCTCTTACCGACCAAGACT 3141
|||||

QY 78 TCTATACCCAGCTCAATAGCGCTCAAGGAGAGA 112
|||||
DB 3140 TCTATACCCAGCTCAATAGCGCTCAAGGAGAGA 3106
|||||

RESULT 10
AX043922/c 349980 bp DNA linear PAT 24-NOV-2000
LOCUS
DEFINITION
Sequence 1 from Patent WO0066791.
ACCESSION
AX043922
VERSION
AX043922.1 GI:11342850
KEYWORDS
Neisseria meningitidis
Neisseria meningitidis
Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
Neisseriaceae; Neisseria.
REFERENCE
1
AUTHORS
Pizza M., Hickey E., Peterson J., Tettelin H., Venter J.C.,
Maignani V., Galeotti C., Mora M., Ratti G., Scarcelli M.,
Scarlato V., Rappuoli R., Frazer C.M. and Grandi G.
Neisseria genomic sequences and methods of their use
Patent: WO 0066791-A 1 09-NOV-2000;
CHIRON CORPORATION (US) ; THE INSTITUTE FOR GENOMIC RESEARCH (US)
LOCATION/Qualifiers
1. 349980
/organism="Neisseria meningitidis"
/mol_type="genomic DNA"
/db_xref="taxon:487"
/note="sequence too long, cut in 8 pieces--seq 1: 1 to
349980 349980 bases--seq 108: 300001 to 649980 349980
bases--seq 109: 600001 to 949980 349980 bases--seq 110:
900001 to 1249980 349980 bases--seq 111: 1200001 to 1549980
349980 bases--seq 112: 1500001 to 1849980 349980 bases--seq
113: 1800001 to 2149980 349980 bases--seq 114: 2100001 to
2272325 172325 bases"
BASE COUNT 83241 a 85091 c 95206 g 86442 t
ORIGIN

Query Match 45.8%; Score 55; DB 6; Length 349980;
Best Local Similarity 81.1%; Pred. No. 2.8e-08;
Matches 77; Conservative 0; Mismatches 15; Indels 3; Gaps 1;

QY 21 AATGAATGCTGTAGGCACGACTGTCTGCGAACGAAACCTCTTACCG---AAGTCT 77
|||||
DB 244794 AATGAATGCTGTAGGCACGACTGTCTGCGAACGAAACCTCTTACCGACCAAGACT 244735
|||||

```

```

QY 78 TCTATACCCAGCTCAATAGCGCTCAAGGAGAGA 112
|||||
DB 244734 TCTATACCCAGCTCAATAGCGCTCAAGGAGAGA 244700
|||||

AC113844 229303 bp DNA linear HTG 19-NOV-2002
Rattus norvegicus clone CH230-134E24, *** SEQUENCING IN PROGRESS
*** 2 unordered pieces.
AC113844
AC113844.6 GI:25072835
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
Rattus norvegicus (Norway rat)
SOURCE
Rattus norvegicus
ORGANISM
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 229303)
AUTHORS
Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaru, A., Barker, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Surrill, K., Caldeira, E.,
Cardenas, V., Carter, K., Cavazos, I., Cesar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, W., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebregregios, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogue, M.,
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorensuewa, L., Loulseghe, H., Lozano, R. J., Lu, X., Ma, J.,
Maheshwari, M., Mahindartine, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhiney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwaokemele, O., Okwono, G., Olarnpunsagoon, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
Plopper, F., Polindexter, A., Popovic, D., Primus, E., Pu, L.,
Fuazo, M., Quito, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J.,
Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajls, D.,
Sheed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
Steinle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,
Weinstock, G. and Gibbs, R. A.
Direct Submission
TITLE
Unpublished
REFERENCE
2 (bases 1 to 229303)
AUTHORS
Worley, K. C.
Direct Submission
TITLE

```

JOURNAL Submitted (05-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE 3 (bases 1 to 229303)

AUTHORS Rat Genome Sequencing Consortium.

TITLE Direct Submission

JOURNAL Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT On Nov 19, 2002 this sequence version replaced gi:23101015. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GSPD  
Center clone name: CH230-134E24  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329  
Consensus quality: 209803 bases at least Q40  
Consensus quality: 212704 bases at least Q30  
Consensus quality: 214539 bases at least Q20  
Estimated insert size: 217292; sum-of-contigs estimation  
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 2 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 114930: contig of 114930 bp in length  
\* 114931 115030: gap of unknown length  
\* 115031 229303: contig of 114273 bp in length.  
Location/Qualifiers  
1..229303  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-134E24"  
1..1020  
/note="wgs\_end\_extension  
clone\_end:Sp6"  
1071..3071  
/note="wgs\_end\_extension  
clone\_end:Sp6"  
clone\_end:Sp6"  
complement(4677..4916)  
/note="clone boundary  
clone\_end:Sp6"  
site:  
end\_sequence: BH348908"  
7547..8335  
/note="clone boundary  
clone\_end:T7"  
site:  
end\_sequence: BH348907"

misc\_feature  
115031..116180  
/note="wgs\_end\_extension  
clone\_end:T7"  
116845..118346  
/note="wgs\_end\_extension  
clone\_end:T7"

misc\_feature  
59433 a 47607 c 48261 g 60802 t 13200 others  
BASE COUNT  
ORIGIN

Query Match 27.3%; Score 32.8; DB 2; Length 229303;  
Best Local Similarity 67.6%; Pred. No. 1.8;  
Matches 46; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 52 CGAAGCAAACTCCTTACCGAAGTCTTATATACCCAGGCTCAATAGCCGCTCAAGGAG 111  
|||||  
Db 14177 CTAACTTACCTCCCTTGAGATGTCCTTACCCAGTCAGATGCTGGCCAGGAGAC 14118  
|||||

QY 112 AGCTATCA 119  
|||||  
Db 14117 AGCCAACA 14110  
|||||

RESULT 12  
BX537269/c 174633 bp DNA linear HTG 10-JUN-2003  
LOCUS Danio rerio clone DKEY-181H1, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 3  
DEFINITION unorderded pieces.  
ACCESSION BX537269  
VERSION BX537269.2 GI:31620796  
KEYWORDS HTG; HTGS PHASE1.  
SOURCE Danio rerio (zebrafish).  
ORGANISM Danio rerio  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;  
Cypriniformes; Cyprinidae; Danio.  
Sims.S.  
1 (bases 1 to 174633)  
REFERENCE  
AUTHORS Direct Submission  
TITLE Submitted (09-JUN-2003) Wellcome Trust Sanger Institute, Hinxton,  
JOURNAL Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
zfish-help@sanger.ac.uk  
COMMENT zfsh-help@sanger.ac.uk  
----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: <http://www.sanger.ac.uk>  
Contact: zfish-help@sanger.ac.uk  
----- Project Information  
Center project name: zk181H1  
----- Summary Statistics  
Assembly program: XGAP4; version 4.5  
Chemistry: Dye-terminator; 100% of reads  
Consensus quality: 173809 bases at least Q40  
Consensus quality: 174028 bases at least Q30  
Consensus quality: 174211 bases at least Q20  
Insert size: 174433; sum-of-contigs  
Insert size: 170374; 4.1% error; agarose-fp  
Quality coverage: 11.23x in Q20 bases; sum-of-contigs Quality  
Coverage: 11.76x in Q20 bases; agarose-fp  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 3 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 9821: contig of 9821 bp in length  
\* 9822 9921: gap of 100 bp  
\* 9922 53645: contig of 43724 bp in length  
\* 53646 53745: gap of 100 bp  
\* 53746 174633: contig of 120888 bp in length.  
Location/Qualifiers

FEATURES  
source

Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, J., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kanat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., MacLean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Medrini, P., Menelus, L., Mihov, I., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R., Vo/A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (22-MAR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Mar 22, 2003 this sequence version replaced gi:28460968.  
All repeats were identified using RepeatMasker:  
Smt, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information  
Center project name: L28903  
Center clone name: 266\_K\_11

----- Summary Statistics  
Sequencing vector: Plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 186175 bases at least Q40  
Consensus quality: 186660 bases at least Q30  
Consensus quality: 186941 bases at least Q20  
Insert size: 193000; agarose-fp  
Insert size: 187586; sum-of-contents  
Quality coverage: 12.1 in Q20 bases; agarose-fp  
Quality coverage: 12.4 in Q20 bases; sum-of-contents

-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 5 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 26328: contig of 26328 bp in length  
26329 26428: gap of 100 bp  
26429 28968: contig of 2540 bp in length  
28969 29068: gap of 100 bp  
29069 145708: contig of 116640 bp in length  
145709 145808: gap of 100 bp  
145809 173362: contig of 27554 bp in length  
173363 173462: gap of 100 bp  
173463 187986: contig of 14524 bp in length.

Location/Qualifiers  
1..187986  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="13"  
/map="13"  
/clone="RP23-266K11"  
/clone\_lib="RPC1-23 Female Mouse BAC"  
1..26328  
/note="assembly\_fragment  
clone end:SP6  
vector side:left"

-----  
nature

```

misc_feature      26429..28968
                    /note="assembly_fragment"
misc_feature      29069..145708
                    /note="assembly_fragment"
misc_feature      145809..173362
                    /note="assembly_fragment"
misc_feature      173463..187986
                    /note="assembly_fragment"
                    clone_end:T7
                    vector_side:right"
BASE COUNT      55004 a 39677 c 38734 g 54171 t 400 others
ORIGIN
Query Match      25.2%; Score 30.2; DB 2; Length 187986;
Best Local Similarity 60.2%; Pred. No. 15;
Matches 50; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
QY 1 CGTTCAGAACAGGCAAGTAAATGCTGAGGACGCGACTGTCTCGAAGCAAAA 60
Db 61038 CAGTCAGGAAAGCAAAATGATGACATGAAACAGCAGCAATCTCTCAAGGAAA 61097
QY 61 ACTCTTACCGAAGTCTTCTATA 83
Db 61098 CATGCAAAACACATGTGATATACA 61120

RESULT 14
AC096226      254098 bp DNA linear HTG 10-MAY-2003
LOCUS      Rattus norvegicus chromosome 18 clone CH230-26M18, WORKING DRAFT
DEFINITION      SEQUENCE, 8 unordered pieces.
ACCESSION      AC096226
VERSION      HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS      Rattus norvegicus
SOURCE      Rattus norvegicus (Norway rat)
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
      Rattus.
REFERENCE      1 (bases 1 to 254098)
AUTHORS      Muzny, D., Marie, Mettaker, M., Lee, Abranzon, S., Adams, C., Alder, J.,
      Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
      Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
      Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
      Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
      Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
      Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
      Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
      Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
      Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
      Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
      Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
      Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
      Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
      Fraser, C. M., Gabisi, A., Gant, R., Garcia, A., Garner, T., Garza, M.,
      Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
      Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
      Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
      Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,
      Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebirt, D., Jackson, A.,
      Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jollivet, A.,
      Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
      Kowis, C., Kratt, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
      Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
      Loresnuhewa, L., Loulseghe, H., Lozano, R. J., Lu, X., Ma, J.,
      Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,
      Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
      Mawney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,
      Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
      Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
      Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
      Nwakoelameh, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,
      Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,

```

```

Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L.,
Fuaro, M., Quiroz, J., Rachlin, E., Reeves, K., Resier, M. A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J.,
Sanders, W., Savery, G., Scherer, S., Scott, G., Shatman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajs, D.,
Sneed, A., Sodargren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
Taylor, I., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
Niederhausen, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,
Weinstock, G., and Gibbs, R. A.
Direct Submission
Unpublished
2 (bases 1 to 254098)
Worley, K. C.
Direct Submission
Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 254098)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GERQ
Center clone name: CH230-26M18
----- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 245379 bases at least Q40
Consensus quality: 246654 bases at least Q30
Consensus quality: 247940 bases at least Q20
Estimated insert size: 258267; sum-of-contigs estimation
Quality coverage: 15x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved
* 1 243470: contig of 243470 bp in length
* 243471 243570: gap of unknown length
* 243571 244710: contig of 1140 bp in length

```

Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, C., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., George, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladus, S.I., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Huiyik, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshewa, L., Louised, H., Lozado, R.J., Lu, X., Ma, J., Maneshwari, M., Mahindartine, M., Mamoud, M., Mallory, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mathew, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Manja, E., Montemayor, J., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Parks, K., Nwackeleneh, O., Okwodu, G., Olarunsgoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, P., Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sison, I., Sitter, C.D., Smajls, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorille, R., Sosa, J., Steidle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Walgren, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlecyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Direct Submission  
Unpublished  
2 (bases 1 to 204173)  
Rat Genome Sequencing Consortium.  
Submitted (13-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One-Baylor Plaza, Houston, TX 77030, USA  
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
----- Project Information  
Center project name: KZKB  
Center clone name: CH230-unknown  
----- Summary Statistics

```

Assembly program: Phrap; version 0.990329
Consensus quality: 185526 bases at least Q40
Consensus quality: 189577 bases at least Q30
Consensus quality: 192542 bases at least Q20
Estimated insert size: 193275; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 202904: contig of 202904 bp in length
* 202905 203004: gap of unknown length
* 203005 204173: contig of 1169 bp in length.
FEATURES
      source
      1..204173
      location/Qualifiers
      1..204173
      /organism="Rattus norvegicus"
      /mol_type="genomic DNA"
      /db_xref="taxon:10116"
      /clone="CH230-unknown"
      misc_feature
      1..11072
      /note="wgs_contig"
BASE COUNT  53905 a 40373 c 40514 g 59594 t 9787 others
ORIGIN
      Query Match      25.0%; Score 30; DB 2; Length 204173;
      Best Local Similarity 54.5%; Pred. No. 18;
      Matches 60; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
      QY 9 AACAGCGAAGGTATGAAATCCCTGAGCGACGGACTGTGTGGAGACGAAACTCCTTA 68
      Db 186037 AAAAAGCTATGTGGTCAAGGTGGTGTGTTGGGCCAGTCTGTCAAGGAAGAGTTTACA 185978
      QY 69 CCGAAGTCTTCTATACCCAGGCTCAATAGCCGCTCAAGGAGAGAGCTATC 118
      Db 185977 CCAAAGTCTTCTAATACGTGGATTGCTTTTCCCAATGGGAGAGTAATTTC 185928

```

Search completed: November 15, 2003, 03:33:48  
 Job time : 536.201 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 14, 2003, 23:55:19 ; Search time 77.4482 Seconds  
(without alignments)  
4182.570 Million cell updates/sec

Title: US-09-928-457-74

Perfect score: 120

Sequence: 1 CGGTCAGAACAGGCAAGGT.....CTCAAGGAGAGAGCTATCAT 120

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseq 19Jun03.\*

- 1: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*
- 2: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*
- 3: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*
- 4: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*
- 5: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*
- 6: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*
- 7: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*
- 8: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*
- 9: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*
- 10: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*
- 11: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*
- 12: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*
- 13: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*
- 14: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*
- 15: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*
- 16: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*
- 17: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*
- 18: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*
- 19: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*
- 20: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*
- 21: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*
- 22: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*
- 23: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*
- 24: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*
- 25: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	120	100.0	120	19 AAV03582	Neisseria meningit
2	120	100.0	120	19 AAV03581	Neisseria meningit
3	120	100.0	120	21 AAA15387	Genomic fragment o
4	120	100.0	120	21 AAA15388	Genomic fragment o
C 5	55.8	46.5	1494	21 AAZ53687	Neisseria meningit
C 6	55.8	46.5	1524	25 ABZ39448	N. gonorrhoeae nuc
C 7	55.8	46.5	1524	25 ABZ41169	N. gonorrhoeae nuc
C 8	55.8	46.5	2094	21 AAA15321	DNA encoding a pol

C 9	55	45.8	1494	21 AAZ53685	Neisseria gonorrhoe
C 10	55	45.8	1494	21 AAZ53686	Neisseria meningit
C 11	55	45.8	2094	21 AAA15333	DNA encoding a pol
C 12	55	45.8	72750	21 AAA81468	N. meningitidis pa
C 13	55	45.8	345980	21 AAF21544	Neisseria meningit
C 14	55	45.8	1437668	21 AAA81490	N. meningitidis B
C 15	28.2	23.5	341511	24 ABS55200	Genomic DNA encodi
C 16	28.2	23.5	506286	24 ABS55320	Human transporter
C 17	28	23.3	2598	24 ABZ13841	Arabidopsis thalia
C 18	28	23.3	2956	21 AAC50837	Arabidopsis thalia
C 19	28	23.3	2957	21 AAC37988	Arabidopsis thalia
C 20	27.2	22.7	464	22 AAS44639	Human full-length
C 21	27.2	22.7	2903	23 ABL29769	Drosophila melanog
C 22	27.2	22.7	5025	23 ABL29768	Drosophila melanog
C 23	27.2	22.7	24417	18 AAT97221	Pseudomonas aerugi
C 24	27.2	22.7	159400	24 ABQ88126	Human osteoblast d
C 25	26.8	22.3	1661	22 ABA20581	Human nervous syst
C 26	26.8	22.3	1661	22 ABA20582	Human nervous syst
C 27	26.8	22.3	2493	22 ABA20580	Human nervous syst
C 28	26.6	22.2	359	21 AAG67192	E. grandis alpha, a
C 29	26.6	22.2	364	21 AAG67194	E. grandis alpha, a
C 30	26.6	22.2	394	25 ABX48531	Bovine EST associa
C 31	26.6	22.2	397	21 AAG67186	E. grandis alpha, a
C 32	26.6	22.2	457	21 AAG67191	E. grandis alpha, a
C 33	26.6	22.2	717	25 ABX08946	CDNA encoding huma
C 34	26.6	22.2	6319	23 ABL18447	Drosophila melanog
C 35	26.6	22.2	29555	23 ABL18446	Drosophila melanog
C 36	26.4	22.0	327	24 ABN76131	Human ORF1078 cDNA
C 37	26.4	22.0	781	21 AAF12631	Aspergillus oryzae
C 38	26.4	22.0	1221	14 AQA46776	Maize dwarf mosaic
C 39	26.4	22.0	2682	22 ABA20583	Human nervous syst
C 40	26.4	22.0	2682	22 AAK84955	Human immune/haema
C 41	26.4	22.0	2687	22 ABA20584	Human nervous syst
C 42	26.4	22.0	2687	22 AAK84957	Human immune/haema
C 43	26.4	22.0	2688	22 ABA20579	Human nervous syst
C 44	26.4	22.0	8543	18 AAT47073	Maize dwarf mosaic
C 45	26.4	22.0	349980	22 AAF86431	Pyrococcus abyssi

ALIGNMENTS

RESULT 1  
AAV03582  
ID AAV03582 standard; DNA, 120 Bp.  
XX AC AAV03582;  
XX DT 22-OCT-1998 (first entry)  
XX DE Neisseria meningitidis DNA sequence C24.  
XX N. gonorrhoeae; N. lactamica; chromosome 22491; region 1; region 2;  
XX region 3; pathogenicity; blood-brain barrier; diagnosis; infection;  
XX meningitis; ss.  
XX OS Neisseria meningitidis.  
XX PN WO9802547-A2.  
XX PD 22-JAN-1998.  
XX PF 11-JUL-1997; 97MO-FR01295.  
XX PR 12-JUL-1996; 96FR-0008768.  
XX PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.  
XX PA (FLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.  
XX PA (SWIK ) SMITHKLINE BEECHAM.  
XX PI Nassif X, Tinsley C, Achtman M, Merker P, Ruelle J,  
XX Vinals C;

DR WPI; 1998-110594/10.  
 XX Genes present in *Neisseria meningitidis* but not other *Neisseria*  
 PT species - and related host cells, RNA, anti-sense sequences,  
 PT polypeptide(s) and antibodies, useful for diagnosing *Neisseria*  
 PT meningitidis infection and in protective vaccines  
 XX Example 4; Page 117; 150pp; French.  
 PS  
 CC AAV03575-606 represent sequences that are present in *Neisseria*  
 CC meningitidis and *N. gonorrhoeae* but not in *N. lactamica*, except for the  
 CC genes involved in biosynthesis of the capsule polysaccharide, *frpA* or *C*,  
 CC *opc*, *porA*, *rotamase*, sequence IC1106, IGA protease, *pilQ*, *pilC*,  
 CC proteins which bind transferrin and opacity proteins. The DNA sequences  
 CC are responsible for the differences in pathogenicity between *N.*  
 CC meningitidis and *N. gonorrhoeae*, specifically they include the genes that  
 CC allow *N. meningitidis* to cross the blood-brain barrier. DNA sequences  
 CC common to *N. meningitidis* and *N. gonorrhoeae*, but absent from *N.*  
 CC *lactamica*, are responsible for colonisation and penetration of the  
 CC mucosa. The DNA sequences can be used to produce probes and primers, and  
 CC antibodies produced against the encoded proteins are used in standard  
 CC hybridisation/immunoassay processes for diagnosis of *N. meningitidis*  
 CC infection, particularly meningitis.  
 XX Sequence 120 BP; 38 A; 30 C; 30 G; 22 T; 0 other;  
 SQ Query Match 100.0%; Score 120; DB 19; Length 120;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-33;  
 Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CGGTGAGAAACAGGCAAGTAAATGAAATGCTGAGGCACGAGCTGTCTCGAACGAAA 60  
 DB 1 CGGTGAGAAACAGGCAAGTAAATGAAATGCTGAGGCACGAGCTGTCTCGAACGAAA 60  
 QY 61 ACTCTTACCGAAGTCTTCTATACCCAGGCTCAATAGCGCTCAAGGAGAGAGCTATCAT 120  
 DB 61 ACTCTTACCGAAGTCTTCTATACCCAGGCTCAATAGCGCTCAAGGAGAGAGCTATCAT 120  
 RESULT 2  
 AAV03581  
 ID AAV03581 standard; DNA; 120 BP.  
 XX  
 AC AAV03581;  
 XX  
 DT 22-OCT-1998 (first entry)  
 DE  
 DE *Neisseria meningitidis* DNA sequence C20.  
 KW *N. gonorrhoeae*; *N. lactamica*; chromosome Z2491; region 1; region 2;  
 KW region 3; pathogenicity; blood-brain barrier; diagnosis; infection;  
 KW meningitis; ss.  
 XX  
 OS *Neisseria meningitidis*.  
 XX WO9802547-A2.  
 XX  
 PD 22-JAN-1998.  
 XX  
 PF 11-JUL-1997; 97WO-FR01295.  
 XX  
 PR 12-JUL-1996; 96PR-0008768.  
 XX  
 PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.  
 PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.  
 PA (SMIK) SMITHKLINE BEECHAM.  
 XX Nassif X, Tinsley C, Achtman M, Merker P, Ruelle J;  
 PI Vinals C;  
 XX WPI; 1998-110594/10.  
 XX Genes present in *Neisseria meningitidis* but not other *Neisseria*

PT species - and related host cells, RNA, anti-sense sequences,  
 PT polypeptide(s) and antibodies, useful for diagnosing *Neisseria*  
 PT meningitidis infection and in protective vaccines  
 XX Example 4; Page 116; 150pp; French.  
 PS  
 CC AAV03575-606 represent sequences that are present in *Neisseria*  
 CC meningitidis and *N. gonorrhoeae* but not in *N. lactamica*, except for the  
 CC genes involved in biosynthesis of the capsule polysaccharide, *frpA* or *C*,  
 CC *opc*, *porA*, *rotamase*, sequence IC1106, IGA protease, *pilQ*, *pilC*,  
 CC proteins which bind transferrin and opacity proteins. The DNA sequences  
 CC are responsible for the differences in pathogenicity between *N.*  
 CC meningitidis and *N. gonorrhoeae*, specifically they include the genes that  
 CC allow *N. meningitidis* to cross the blood-brain barrier. DNA sequences  
 CC common to *N. meningitidis* and *N. gonorrhoeae*, but absent from *N.*  
 CC *lactamica*, are responsible for colonisation and penetration of the  
 CC mucosa. The DNA sequences can be used to produce probes and primers, and  
 CC antibodies produced against the encoded proteins are used in standard  
 CC hybridisation/immunoassay processes for diagnosis of *N. meningitidis*  
 CC infection, particularly meningitis.  
 XX Sequence 120 BP; 38 A; 30 C; 30 G; 22 T; 0 other;  
 SQ Query Match 100.0%; Score 120; DB 19; Length 120;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-33;  
 Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CGGTGAGAAACAGGCAAGTAAATGAAATGCTGAGGCACGAGCTGTCTCGAACGAAA 60  
 DB 1 CGGTGAGAAACAGGCAAGTAAATGAAATGCTGAGGCACGAGCTGTCTCGAACGAAA 60  
 QY 61 ACTCTTACCGAAGTCTTCTATACCCAGGCTCAATAGCGCTCAAGGAGAGAGCTATCAT 120  
 DB 61 ACTCTTACCGAAGTCTTCTATACCCAGGCTCAATAGCGCTCAAGGAGAGAGCTATCAT 120  
 RESULT 3  
 AAAL5387  
 ID AAAL5387 standard; DNA; 120 BP.  
 XX  
 AC AAAL5387;  
 XX  
 DT 04-SEP-2000 (first entry)  
 DE  
 DE Genomic fragment of *Neisseria meningitidis* Z2491.  
 KW Pathogenic strain; *Neisseria*; vaccine; *Neisseria* infection; ss.  
 XX  
 OS *Neisseria meningitidis*.  
 XX WO200026375-A2.  
 XX  
 PD 11-MAY-2000.  
 XX  
 PF 28-OCT-1999; 99WO-FR02643.  
 XX  
 PR 30-OCT-1998; 98PR-0013693.  
 XX  
 PA (INRM) PASTEUR MERIEUX SERUMS & VACCINS SA.  
 PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.  
 XX Aujame L, Bouchardon A, Renauld-Mongenie G, Rokbi B, Nassif X;  
 PI Tinsley C, Perrin A;  
 XX WPI; 2000-365622/31.  
 XX New polypeptide specific for pathogenic *Neisseria* useful in therapeutic  
 PT or preventative vaccines and for diagnosis  
 XX Claim 1; Page 39; 187pp; French.  
 XX The present sequence represents a genomic fragment of *Neisseria*  
 CC meningitidis. The specification describes proteins that are specific



CC for pathogenic strains of *Neisseria*. The polynucleotides, polypeptides,  
 CC or their antigenic fragments, are used in vaccines to treat or protect  
 CC against *Neisseria* infections, particularly by *N. meningitidis*. The  
 CC polynucleotide sequences are also used for recombinant production of  
 CC the polypeptide and to produce attenuated *Neisseria* strains that  
 CC overexpress it, or express it in a non-toxic mutant form.  
 CC note: the present sequence is included in Claim 1, but it is those  
 CC sequences that do not include the present sequence that are actually  
 CC claimed.

XX SQ Sequence 120 BP; 38 A; 30 C; 30 G; 22 T; 0 other;

Query Match 100.0%; Score 120; DB 21; Length 120;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-33;  
 Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGTGAGAAACAGCGAAGTAAATGAAATGCTTACCGGCTCAATAGCGGCTCAAGGAGAGACTATCAT 60  
 Db 1 CGGTGAGAAACAGCGAAGTAAATGAAATGCTTACCGGCTCAATAGCGGCTCAAGGAGAGACTATCAT 60  
 QY 61 ACTCCTTACCGAAGTCTTCTATACCGAGGCTCAATAGCGGCTCAAGGAGAGACTATCAT 120  
 Db 61 ACTCCTTACCGAAGTCTTCTATACCGAGGCTCAATAGCGGCTCAAGGAGAGACTATCAT 120

RESULT 4

AAAL5388  
 ID AAA15388 standard; DNA; 120 BP.

AC AAA15388;

DT 04-SEP-2000 (first entry)

DE Genomic fragment of *Neisseria meningitidis* 22491.

KW Pathogenic strain; *Neisseria*; vaccine; *Neisseria* infection; ss.

OS *Neisseria meningitidis*.

PN WO200026375-A2.

XX 11-MAY-2000.

PF 28-OCT-1999; 99WO-FR02643.

XX 30-OCT-1998; 98FR-0013693.

PA (INRM) PASTEUR MERIEUX SERUMS & VACCINS SA.

XX (INRM) INSERM INST NAT SANTE & RECH MEDICALE.

PI Aujame L, Bouchardon A, Renauld-Mongenie G, Rokhi B, Nassif X;

PI Tinsley C, Perrin A;

XX WPI; 2000-365622/31.

PT New polypeptide specific for pathogenic *Neisseria* useful in therapeutic  
 or preventative vaccines and for diagnosis

PS Claim 1; Page 39; 187pp; French.

XX The present sequence represents a genomic fragment of *Neisseria*  
 CC meningitidis. The specification describes proteins that are specific  
 CC for pathogenic strains of *Neisseria*. The polynucleotides, polypeptides,  
 CC or their antigenic fragments, are used in vaccines to treat or protect  
 CC against *Neisseria* infections, particularly by *N. meningitidis*. The  
 CC polynucleotide sequences are also used for recombinant production of  
 CC the polypeptide and to produce attenuated *Neisseria* strains that  
 CC overexpress it, or express it in a non-toxic mutant form.  
 CC note: the present sequence is included in Claim 1, but it is those  
 CC sequences that do not include the present sequence that are actually  
 CC claimed.

XX SQ Sequence 120 BP; 38 A; 30 C; 30 G; 22 T; 0 other;

Query Match 100.0%; Score 120; DB 21; Length 120;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-33;  
 Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGTGAGAAACAGCGAAGTAAATGAAATGCTTACCGGCTCAATAGCGGCTCAAGGAGAGACTATCAT 60  
 Db 1 CGGTGAGAAACAGCGAAGTAAATGAAATGCTTACCGGCTCAATAGCGGCTCAAGGAGAGACTATCAT 60

QY 61 ACTCCTTACCGAAGTCTTCTATACCGAGGCTCAATAGCGGCTCAAGGAGAGACTATCAT 120  
 Db 61 ACTCCTTACCGAAGTCTTCTATACCGAGGCTCAATAGCGGCTCAAGGAGAGACTATCAT 120

RESULT 5

AAZ53687/C

ID AAZ53687 standard; DNA; 1494 BP.

XX AAZ53687;

DT 21-MAR-2000 (first entry)

DE *Neisseria meningitidis* ORF 402 partial DNA sequence SEQ ID NO:1323.

XX *Neisseria meningitidis*; *Neisseria gonorrhoeae*; antigen; vaccine;  
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;  
 KW antibacterial; gene therapy; ds.

OS *Neisseria meningitidis*.

PN WO9957280-A2.

XX 11-NOV-1999.

PF 30-APR-1999; 99WO-US09346.

XX 01-MAY-1998; 98US-0083758.

PR 31-JUL-1998; 98US-0094869.

PR 02-SEP-1998; 98US-0098994.

PR 02-SEP-1998; 98US-0099062.

PR 09-OCT-1998; 98US-0103749.

PR 09-OCT-1998; 98US-0103794.

PR 09-OCT-1998; 98US-0103796.

PR 25-FEB-1999; 99US-0121528.

XX (CHIR) CHIRON CORP.

PA (GENO-) INST GENOMIC RES.

XX Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;

PI Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;

PI Tettelin H, Venter JC;

XX WPI; 2000-062150/05.

DR P-PSDB; AAY74925.

PT Novel *Neisseria* polypeptides predicted to be useful antigens for  
 vaccines and diagnostics

PS Claim 7; Page 726; 1453pp; English.

XX AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941  
 CC represent novel *Neisseria meningitidis* and *N. gonorrhoeae* polynucleotides  
 CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent  
 CC PCR primers used in the exemplification of the present invention. The  
 CC polypeptides, the polynucleotides, antibodies and compositions of  
 CC the invention can be used as vaccines, as diagnostic reagents, and as  
 CC immunogenic compositions. The polypeptides can be used in the  
 CC manufacture of medicaments for treating or preventing infection due to  
 CC *Neisseria* bacteria (e.g. meningitis and septicaemia), to detect the  
 CC presence of *Neisseria* bacteria, or to raise antibodies. They may also  
 CC be used to screen for agonists or antagonists, which may themselves  
 CC have use as antibacterial agents. The polynucleotides of the invention  
 CC may also be used in gene therapy protocols.

XX SQ Sequence 1494 BP; 321 A; 354 C; 373 G; 446 T; 0 other;  
Query Match 46.5%; Score 55.8; DB 21; Length 1494;  
Best Local Similarity 79.8%; Pred. No. 5.4e-10;  
Matches 79; Conservative 0; Mismatches 17; Indels 3; Gaps 1;  
QY 17 AGTAAATGAAATGCTGAGCAGCGACTGTGCTGCGAACGAAATCTCTTACCG---AA 73  
DB 147 AGTAAATGAAATGCTGAGCAGCGACTGTGCTGCGAACGAAATCTCTTACCGCAA 88  
QY 74 GTCTTCTATACCCAGGCTCAATAGCGCTCAAGGAGAGA 112  
DB 87 GACTTCTATACCCAGGCTCAATAGCGCTCAAGGAGAGA 49  
RESULT 6  
ABZ39448  
ID ABZ39448 standard; DNA; 1524 BP.  
XX AC ABZ39448;  
XX DT 07-MAR-2003 (first entry)  
XX DE N. gonorrhoeae nucleotide sequence SEQ ID 3485.  
XX KW Antibacterial; infection; vaccine; gene therapy; gene; ds.  
XX OS Neisseria gonorrhoeae.  
XX PN WO200279243-A2.  
XX PD 10-OCT-2002.  
XX PF 12-FEB-2002; 2002WO-IB02069.  
XX PR 12-FEB-2001; 2001GB-0003424.  
XX PA (CHIR-) CHIRON SPA.  
XX PI Fontana MR, Pizsa M, Massignani V, Monaci E;  
XX DR WPI; 2003-058415/05.  
XX DR P-PSDB; ABP78478.  
XX PT New protein from Neisseria gonorrhoeae, useful for the manufacture of a  
XX PS medicament for treating or preventing N. gonorrhoeae infection -  
XX Disclosure; Page 443-444; 815pp; English.  
XX The present invention relates to proteins from Neisseria gonorrhoeae.  
XX Also disclosed are the nucleic acid molecules encoding the proteins and  
XX antibodies that specifically bind to the proteins. The composition  
XX comprising the protein, nucleic acid or antibody is useful for the  
XX manufacture of a medicament for treating or preventing N. gonorrhoeae  
XX infection, this may be in the form of a vaccine or gene therapy.  
XX Sequences given in records ABZ37706-ABZ42016 represent nucleic acid  
XX molecules of the invention.  
XX SQ Sequence 1524 BP; 458 A; 386 C; 355 G; 325 T; 0 other;  
Query Match 46.5%; Score 55.8; DB 25; Length 1524;  
Best Local Similarity 79.8%; Pred. No. 5.4e-10;  
Matches 79; Conservative 0; Mismatches 17; Indels 3; Gaps 1;  
QY 17 AGTAAATGAAATGCTGAGCAGCGACTGTGCTGCGAACGAAATCTCTTACCG---AA 73  
DB 1345 AGTAAATGAAATGCTGAGCAGCGACTGTGCTGCGAACGAAATCTCTTACCGCAA 1404  
QY 74 GTCTTCTATACCCAGGCTCAATAGCGCTCAAGGAGAGA 112  
DB 1405 GACTTCTATACCCAGGCTCAATAGCGCTCAAGGAGAGA 1443

RESULT 7  
ABZ41169/c  
ID ABZ41169 standard; DNA; 1524 BP.  
XX AC ABZ41169;  
XX DT 07-MAR-2003 (first entry)  
XX DE N. gonorrhoeae nucleotide sequence SEQ ID 6927.  
XX KW Antibacterial; infection; vaccine; gene therapy; gene; ds.  
XX OS Neisseria gonorrhoeae.  
XX PN WO200279243-A2.  
XX PD 10-OCT-2002.  
XX PF 12-FEB-2002; 2002WO-IB02069.  
XX PR 12-FEB-2001; 2001GB-0003424.  
XX PA (CHIR-) CHIRON SPA.  
XX PI Fontana MR, Pizsa M, Massignani V, Monaci E;  
XX DR WPI; 2003-058415/05.  
XX DR P-PSDB; ABP80199.  
XX PT New protein from Neisseria gonorrhoeae, useful for the manufacture of a  
XX PS medicament for treating or preventing N. gonorrhoeae infection -  
XX Disclosure; Page 690; 815pp; English.  
XX The present invention relates to proteins from Neisseria gonorrhoeae.  
XX Also disclosed are the nucleic acid molecules encoding the proteins and  
XX antibodies that specifically bind to the proteins. The composition  
XX comprising the protein, nucleic acid or antibody is useful for the  
XX manufacture of a medicament for treating or preventing N. gonorrhoeae  
XX infection, this may be in the form of a vaccine or gene therapy.  
XX Sequences given in records ABZ37706-ABZ42016 represent nucleic acid  
XX molecules of the invention.  
XX SQ Sequence 1524 BP; 325 A; 355 C; 386 G; 458 T; 0 other;  
Query Match 46.5%; Score 55.8; DB 25; Length 1524;  
Best Local Similarity 79.8%; Pred. No. 5.4e-10;  
Matches 79; Conservative 0; Mismatches 17; Indels 3; Gaps 1;  
QY 17 AGTAAATGAAATGCTGAGCAGCGACTGTGCTGCGAACGAAATCTCTTACCG---AA 73  
DB 180 AGTAAATGAAATGCTGAGCAGCGACTGTGCTGCGAACGAAATCTCTTACCGCAA 121  
QY 74 GTCTTCTATACCCAGGCTCAATAGCGCTCAAGGAGAGA 112  
DB 120 GACTTCTATACCCAGGCTCAATAGCGCTCAAGGAGAGA 82  
RESULT 8  
AAA15321/c  
ID AAA15321 standard; DNA; 2094 BP.  
XX AC AAA15321;  
XX DT 04-SEP-2000 (first entry)  
XX DE DNA encoding a polypeptide of a Neisseria pathogenic strain.  
XX KW Pathogenic strain; Neisseria; vaccine; Neisseria infection; ss.  
XX OS Neisseria meningitidis.  
XX

Key	Location/Qualifiers
FT	1..2094
FT	/*tag= a
XX	
XX	WO200026375-A2.
XX	
XX	11-MAY-2000.
XX	
XX	28-OCT-1999; 99WO-FR02643.
XX	
XX	30-OCT-1998; 98FR-0013693.
XX	
XX	(INNR ) PASTEUR MERIEUX SERUMS & VACCINS SA.
XX	(INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX	
XX	Auajme L, Bouchardon A, Renauld-Mongenie G, Rokbi B, Nassif X;
XX	Tinsley C, Perrin A;
XX	
XX	WPI; 2000-365622/31.
XX	P-PSDB; AAY93290.
XX	
XX	New polypeptide specific for pathogenic Neisseria useful in therapeutic
XX	or preventative vaccines and for diagnosis -
XX	
XX	Claim 2; Page 125-128; 187pp; French.
XX	
XX	The present sequence encodes a protein that is specific for pathogenic
XX	strains of Neisseria. the polynucleotides, polypeptides, or their
XX	antigenic fragments, are used in vaccines to treat or protect against
XX	Neisseria infections, particularly by N. meningitidis. The
XX	polynucleotide sequence is also used for recombinant production of
XX	the polypeptide and to produce attenuated Neisseria strains that
XX	overexpress it, or express it in a non-toxic mutant form.
XX	
XX	Sequence 2094 BP; 428 A; 500 C; 512 G; 654 T; 0 other;
XX	
XX	Query Match 46.5%; Score 55.8; DB 21; Length 2094;
XX	Best Local Similarity 79.8%; Pred. No. 6e-10;
XX	Matches 79; Conservative 0; Mismatches 17; Indels 3; Gaps 1;
XX	
XX	17 AGTAAATCAAAATGCTGAGCAGGACTGTGCTGGACGAAACTCTTACCG---AA 73
XX	747 AGTAAATCAAAATGCTGAGCAGGACTGTGCTGGACGAAACTCTTACCGCAA 688
XX	
XX	74 GTCTTCTATACCCAGGCTCAATAGCCGCTCAAGGAGAGA 112
XX	687 GACTTCTATACCAAGCTCAATAGCCGCTAAGGAAGA 649
XX	
XX	RESULT 9
XX	AAZ53685/c
XX	ID AAZ53685 standard; DNA; 1494 BP.
XX	XX AAZ53685;
XX	XX AC
XX	XX AC
XX	XX
XX	21-MAR-2000 (first entry)
XX	
XX	Neisseria gonorrhoeae ORF 402 partial DNA sequence SEQ ID NO:1319.
XX	
XX	Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
XX	antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
XX	antibacterial; gene therapy; ds.
XX	
XX	Neisseria gonorrhoeae.
XX	
XX	WO9957280-A2.
XX	
XX	11-NOV-1999.
XX	
XX	30-APR-1999; 99WO-US09346.
XX	
XX	01-MAY-1998; 98US-0083758.
XX	31-JUL-1998; 98US-0094869.

```

PR 01-MAY-1998; 98US-0083758.
PR 21-JUL-1998; 98US-0094869.
PR 02-SEP-1998; 98US-0098994.
PR 02-SEP-1998; 98US-0099062.
PR 09-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103794.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.
XX (CHIR ) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
XX Fraser C, Galeotti C, Grandi G, Hickey E, Maignani V, Mora M;
XX Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
XX Tettelin H, Venter JC;
XX WPI; 2000-062150/05.
DR P-PSDB; AAY74924.
XX
XX Novel Neisserial polypeptides predicted to be useful antigens for
XX vaccines and diagnostics
XX
XX Claim 7; Page 724-725; 1453pp; English.
XX
XX AA253015 to AA254536, AA254577 to AA254615, and AAY74253 to AAY75941
XX represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
XX and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent
XX PCR primers used in the exemplification of the present invention. The
XX polypeptides, the polynucleotides, antibodies and compositions of
XX the invention can be used as vaccines, as diagnostic reagents, and as
XX immunogenic compositions. The polypeptides can be used in the
XX manufacture of medicaments for treating or preventing infection due to
XX Neisseria meningitidis (e.g. meningitis and septicaemia), to detect the
XX presence of Neisseria meningitidis bacteria, or to raise antibodies. They may also
XX be used to screen for agonists or antagonists, which may themselves
XX have use as antibacterial agents. The polynucleotides of the invention
XX may also be used in gene therapy protocols.
XX
XX Sequence 1494 BP; 319 A; 354 C; 373 G; 442 T; 6 other;
SQ
XX
XX Query Match 45.8%; Score 55; DB 21; Length 1494;
XX Best Local Similarity 81.1%; Pred. No. 1e-09;
XX Matches 77; Conservative 15; Mismatches 15; Indels 3; Gaps 1;
XX
XX 21 AATGAAATGCTGAGGACGAGCTGTGCTGCGAACGAAACTCTTACCG---AAGTCT 77
XX |||||
XX 143 AATGAAATGCTGAGGACGAGCTGTGCTGCGAACGAAACTCTTACCGCAAGACT 84
XX
XX 78 TCTATACCCAGCTCAATAGCCGCTCAAGGAGAGA 112
XX |||||
XX 83 TCTATACCCAGCTCAATAGCCGCTTAGGAAAGA 49
XX
XX RESULT 11
XX AA15333/c
XX ID AA15333 standard; DNA; 2094 BP.
XX
XX AC AA15333;
XX
XX 04-SEP-2000 (first entry)
XX
XX DNA encoding a polypeptide of a Neisseria pathogenic strain.
XX
XX Pathogenic strain; Neisseria; vaccine; Neisseria infection; ss.
XX
XX Neisseria gonorrhoeae.
XX
XX Key Location/Qualifiers
XX CDS 1..2094
XX /*tag= a
XX
XX WO200026375-A2.
XX

```

```

PD 11-MAY-2000.
XX
XX 28-OCT-1999; 99WO-FR02643.
XX
XX 30-OCT-1998; 98FR-0013693.
XX
XX (INMR ) PASTEUR MERIEUX SERUMS & VACCINS SA.
XX (INMR ) INSERM INST NAT SANTE & RECH MEDICALE.
XX
XX Aujame L, Bouchardon A, Renauld-Mongenie G, Rokbi B, Nassif X;
XX Tinsley C, Perrin A;
XX
XX WPI; 2000-365622/31.
XX P-PSDB; AAY93303.
XX
XX New polypeptide specific for pathogenic Neisseria useful in therapeutic
XX or preventative vaccines and for diagnosis
XX
XX Claim 4; Page 170-173; 187pp; French.
XX
XX The present sequence encodes a protein that is specific for pathogenic
XX strains of Neisseria. The polynucleotides, polypeptides, or their
XX antigenic fragments, are used in vaccines to treat or protect against
XX Neisseria infections, particularly by N. meningitidis. The
XX polynucleotide sequence is also used for recombinant production of
XX the polypeptide and to produce attenuated Neisseria strains that
XX overexpress it, or express it in a non-toxic mutant form.
XX
XX Sequence 2094 BP; 428 A; 495 C; 515 G; 556 T; 0 other;
SQ
XX
XX Query Match 45.8%; Score 55; DB 21; Length 2094;
XX Best Local Similarity 81.1%; Pred. No. 1.2e-09;
XX Matches 77; Conservative 15; Mismatches 15; Indels 3; Gaps 1;
XX
XX 21 AATGAAATGCTGAGGACGAGCTGTGCTGCGAACGAAACTCTTACCG---AAGTCT 77
XX |||||
XX 743 AATGAAATGCTGAGGACGAGCTGTGCTGCGAACGAAACTCTTACCGCAAGACT 684
XX
XX 78 TCTATACCCAGCTCAATAGCCGCTCAAGGAGAGA 112
XX |||||
XX 83 TCTATACCCAGCTCAATAGCCGCTTAGGAAAGA 649
XX
XX RESULT 12
XX AA81468
XX ID AA81468 standard; DNA; 72750 BP.
XX
XX AC AA81468;
XX
XX 04-DEC-2000 (first entry)
XX
XX N. meningitidis partial DNA sequence gnm_16 SEQ ID NO:16.
XX
XX Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
XX antigen; vaccine; diagnosis; infection; antibacterial; identification;
XX Meningococcus B; MenB; ds.
XX
XX Neisseria meningitidis.
XX
XX WO200022430-A2.
XX
XX 20-APR-2000.
XX
XX 08-OCT-1999; 99WO-US23573.
XX
XX 09-OCT-1998; 98US-0103794.
XX 30-APR-1999; 99US-0132068.
XX
XX (CHIR ) CHIRON CORP.
XX
XX Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;
XX Maignani V, Galeotti C, Mora M, Ratti G, Scarselli M,
XX Rappuoli R, Pizza M;

```

XX WPI; 2000-318079/27.  
 XX  
 XX Isolated nucleotide sequences of *Neisseria meningitidis* which can be  
 PT used in the diagnosis and treatment of *N. meningitidis* infection and  
 PT other *Neisseria* infections, for example, *N. gonorrhoea* -  
 XX  
 PS Claim 7; Page 406-428; 1760pp; English.  
 XX  
 XX The present invention describes methods of obtaining immunogenic  
 CC proteins from *Neisseria* genomic sequences. AA81453 to AA82414  
 CC represent specifically claimed *Neisseria meningitidis* genomic DNA  
 CC sequences; AA82460 to AA8303 and AA82520 to AA85683 represent  
 CC *Neisseria* DNA sequences and their corresponding proteins; AA81254 to  
 CC AA81259 and AA81304 to AA81321 represent PCR primers used in the  
 CC isolation of *Neisseria meningitidis* DNA sequences; and AA81322 to  
 CC AA81452 represent *Neisseria meningitidis* MenB polynucleotide ORF  
 CC sequences, which are all used in the exemplification of the present  
 CC invention. The nucleic acid sequences, protein sequences, and antibodies  
 CC against them can be used in the manufacture of a composition. The  
 CC composition can be used as a medicament (or in the manufacture of a  
 CC medicament) for treating, preventing or diagnosing infection due to  
 CC *Neisseria* bacteria. For example, some of the identified proteins could  
 CC be components of vaccines against *Meningococcus B*; against all serotypes;  
 CC and/or against all pathogenic *Neisseriae*. Identification of sequences  
 CC from the bacterium will also facilitate production of biological probes,  
 CC particularly organism-specific probes. Attempts to make efficacious  
 CC *Meningococcus B* vaccines have failed mainly due to antigen tolerance.  
 CC Multivalent vaccines have also been tried but none have successfully  
 CC overcome antigenic variability. The provision of further, complete  
 CC sequences may provide an opportunity to identify secreted or surface  
 CC exposed proteins that may be presumed targets for the immune system and  
 CC which are not antigenically variable or at least more conserved than  
 CC other more variable regions.  
 XX  
 XX Sequence 72750 BP; 17518 A; 19945 C; 18810 G; 16477 T; 0 other;  
 SQ  
 Query Match 45.8%; Score 55; DB 21; Length 72750;  
 Best Local Similarity 81.1%; Pred. No. 4e-09;  
 Matches 77; Conservative 0; Mismatches 15; Indels 3; Gaps 1;  
 QY 21 AATGAAATGCTGAGGCACGGACTGTGCGAAGCAAACTCCTTACCG---AAGTCT 77  
 DB 56066 AATGAAATGCTGAGGCACGGACTGTGCGAAGCAAACTCCTCACCACAGACT 56125  
 QY 78 TCTATACCCAGGCTCAATAGCCGCTCAAGGAGAGA 112  
 DB 56126 TCTATACCCAGGCTCAATAGCCGCTCAAGGAGAGA 56160  
 RESULT 13  
 AAF21544/c  
 ID AAF21544 standard; DNA; 349980 BP.  
 XX AAF21544;  
 XX  
 XX 13-MAR-2001 (first entry)  
 DT  
 DE *Neisseria meningitidis B* nucleotide sequence SEQ ID NO:1.  
 XX  
 XX *Neisseria meningitidis*; *Neisseria gonorrhoeae*; immunogenic; vaccine;  
 KW diagnosis; antigen; detection; infection; gene therapy; antibacterial;  
 KW ds.  
 XX  
 XX *Neisseria meningitidis*.  
 OS  
 XX  
 XX WO2000066791-A1.  
 PN  
 XX  
 XX 09-NOV-2000.  
 PD  
 XX  
 XX 08-MAR-2000; 2000WO-US05928.  
 PF  
 XX  
 XX 30-APR-1999; 99US-0132068.  
 PR

08-OCT-1999; 99WO-US23573.  
 28-FEB-2000; 2000GB-0004695.  
 (CHIR ) CHIRON CORP.  
 (GENO-) INST GENOMIC RES.  
 Piza M, Hickey E, Peterson J, Tettelin H, Venter JC, Masignani V,  
 Galeotti C, Mora M, Ratti G, Scarselli M, Scariato V, Rappuoli R;  
 Frazer CM, Grandi G;  
 WPI; 2000-647603/62.  
 XX  
 XX *Neisseria meningitidis B* full length genome sequence and open reading  
 PT frames are used to detect, treat and prevent *Neisseria* infections -  
 XX  
 PS Claim 7; Appendix A; 692pp; English.  
 XX  
 XX The present invention describes the full length genome of  
 CC *Neisseria meningitidis B* (NMB). The sequences in AAF21544 and AAF21607  
 CC to AAF21613 represent fragments of the NMB genomic sequence, as the  
 CC sequence was too long to go in a record on its own it was split into 8  
 CC sequences which overlap each other at the beginning and end of each  
 CC sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at  
 CC the beginning of AAF21607, the last 49980 bp of AAF21607 are repeated at  
 CC the beginning of AAF21608, and so on). AAF21545 to AAF21598 encode the  
 CC *Neisseria* proteins given in AAB58550 to AAB58593, and AAF21599 to  
 CC AAF21606 represent PCR primers which are used in the exemplification of  
 CC the present invention. The NMB genome and fragments from it have  
 CC antibacterial activity, and can be used in vaccines and gene therapy.  
 CC *Neisseria* nucleic acids, proteins and/or antibodies which binds to the  
 CC proteins can be used in compositions for treating or preventing infection  
 CC due to *Neisseria* bacteria or as a diagnostic reagent for detecting the  
 CC presence of *Neisseria* bacteria or of antibodies raised to *Neisseria*  
 CC bacteria. Computers, computer memory, computer storage medium or computer  
 CC databases can be used in a search to identify open reading frames (ORFs)  
 CC or coding sequences within the NMB genome. The DNA sequences provide  
 CC further opportunities to find antigenic or immunogenic proteins which are  
 CC more effective in vaccines than the outer membrane proteins currently  
 CC used.  
 XX  
 XX Sequence 349980 BP; 83241 A; 85091 C; 95206 G; 86442 T; 0 other;  
 SQ  
 Query Match 45.8%; Score 55; DB 21; Length 349980;  
 Best Local Similarity 81.1%; Pred. No. 7e-09;  
 Matches 77; Conservative 0; Mismatches 15; Indels 3; Gaps 1;  
 QY 21 AATGAAATGCTGAGGCACGGACTGTGCGAAGCAAACTCCTTACCG---AAGTCT 77  
 DB 244794 AATGAAATGCTGAGGCACGGACTGTGCGAAGCAAACTCCTCACCACAGACT 244735  
 QY 78 TCTATACCCAGGCTCAATAGCCGCTCAAGGAGAGA 112  
 DB 244734 TCTATACCCAGGCTCAATAGCCGCTCAAGGAGAGA 244700  
 RESULT 14  
 AAF21490/c  
 ID AAF21490 standard; DNA; 1437668 BP.  
 XX AAF21490;  
 AC  
 XX  
 XX 04-DEC-2000 (first entry)  
 DT  
 DE *N. meningitidis B* full length genome DNA sequence SEQ ID NO:1068.  
 XX  
 XX *Neisseria meningitidis*; *Neisseria gonorrhoeae*; genome; immunogenic;  
 KW antigen; vaccine; diagnosis; infection; antibacterial; identification;  
 KW *Meningococcus B*; MenB; ds.  
 XX  
 XX *Neisseria meningitidis*.  
 OS  
 XX  
 XX WO200022430-A2.  
 PN  
 XX

PD 20-APR-2000.  
XX XX  
PF 08-OCT-1999; 99WO-US23573.  
XX XX  
PR 09-OCT-1998; 98US-0103794.  
PR 30-APR-1999; 99US-0132068.  
XX XX  
XX (CHIR ) CHIRON CORP.  
PA  
PI Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;  
PI Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;  
PI Rappuoli R, Pizza M;  
XX  
DR WPI; 2000-318079/27.  
XX  
XX Isolated nucleotide sequences of *Neisseria meningitidis* which can be  
PT used in the diagnosis and treatment of *N. meningitidis* infection and  
PT other *Neisseria* infections, for example, *N. gonorrhoea*.  
XX  
PS Claim 7; Page 866-1272; 1760pp; English.  
XX  
CC The present invention describes methods of obtaining immunogenic  
CC proteins from *Neisseria* genomic sequences. AA81453 to AA82414  
CC represent specifically claimed *Neisseria meningitidis* genomic DNA  
CC sequences; AA81260 to AA81303 and AA82562 to AA82563 represent  
CC *Neisseria* DNA sequences and their corresponding proteins; AA81254 to  
CC AA81259 and AA81304 to AA81321 represent PCR primers used in the  
CC isolation of *Neisseria meningitidis* DNA sequences; and AA81322 to  
CC AA81452 represent *Neisseria meningitidis* MenB polynucleotide ORF  
CC sequences, which are all used in the exemplification of the present  
CC invention. The nucleic acid sequences, protein sequences, and antibodies  
CC against them, can be used in the manufacture of a composition. The  
CC composition can be used as a medicament (or in the manufacture of a  
CC medicament) for treating, preventing or diagnosing infection due to a  
CC *Neisseria* bacteria. For example, some of the identified proteins could  
CC be components of vaccines against *Meningococcus B*; against all serotypes;  
CC and/or against all pathogenic *Neisseriae*. Identification of sequences  
CC from the bacterium will also facilitate production of biological probes,  
CC particularly organism-specific probes. Attempts to make efficacious  
CC *Meningococcus B* vaccines have failed mainly due to antigen tolerance.  
CC Multivalent vaccines have also been tried but none have successfully  
CC overcome antigenic variability. The provision of further, complete  
CC sequences may provide an opportunity to identify secreted or surface  
CC exposed proteins that may be presumed targets for the immune system and  
CC which are not antigenically variable or at least more conserved than  
CC other more variable regions.  
XX  
SQ Sequence 1437668 BP; 344338 A; 353206 C; 385074 G; 355045 T; 5 other;  
Query Match 45.8%; Score 55; DB 21; Length 1437668;  
Best Local Similarity 81.1%; Pred. No. 1.1e-08;  
Matches 77; Conservative 0; Mismatches 15; Indels 3; Gaps 1;  
QY 21 AATGAATGCTGAGCAGCGACTGTGCTGCCAAGCAAACTCCCTACCG---AGTCT 77  
Db 244794 AATGAATGCTGAGCAGCGACTGTGCTGCCAAGCAAACTCCCTACCGCAAGACT 244735  
QY 78 TCTATACCCAGGCTCAATAGCCGCTCAAGGAGAGA 112  
Db 244734 TCTATACCCAGGCTCAATAGCCGCTCAAGGAGAGA 244700  
RESULT 15  
ABS55200  
ID ABS55200 standard; DNA; 341511 BP.  
XX AC ABS55200;  
XX AC  
XX DT 06-JAN-2003 (first entry)  
XX DE Genomic DNA encoding human transporter protein.  
XX KW Human; transporter protein; hypoglycaemia; antidiabetic; gene therapy;

gene; ds; single nucleotide polymorphism; SNP.  
XX XX Homo sapiens.  
XX XX Location/Qualifiers  
XX XX 2001..339512  
XX XX /tag= a  
XX XX /products= "Human transporter protein"  
XX XX 2001..2073  
XX FT exon  
XX FT /tag= b  
XX FT /number= 1  
XX FT 2074..30237  
XX FT intron  
XX FT /tag= c  
XX FT /number= 1  
XX FT 30238..30361  
XX FT exon  
XX FT /tag= d  
XX FT /number= 2  
XX FT 30362..40719  
XX FT intron  
XX FT /tag= e  
XX FT /number= 2  
XX FT 40720..40826  
XX FT exon  
XX FT /tag= f  
XX FT /number= 3  
XX FT 40827..45732  
XX FT intron  
XX FT /tag= g  
XX FT /number= 3  
XX FT 45733..45861  
XX FT exon  
XX FT /tag= h  
XX FT /number= 4  
XX FT 45862..59818  
XX FT intron  
XX FT /tag= i  
XX FT /number= 4  
XX FT 59819..59934  
XX FT exon  
XX FT /tag= j  
XX FT /number= 5  
XX FT 59935..65742  
XX FT intron  
XX FT /tag= k  
XX FT /number= 5  
XX FT 65743..66135  
XX FT exon  
XX FT /tag= l  
XX FT /number= 6  
XX FT 66136..96458  
XX FT intron  
XX FT /tag= m  
XX FT /number= 6  
XX FT 96459..96885  
XX FT exon  
XX FT /tag= n  
XX FT /number= 7  
XX FT 96886..197167  
XX FT intron  
XX FT /tag= o  
XX FT /number= 7  
XX FT 197168..197367  
XX FT exon  
XX FT /tag= p  
XX FT /number= 8  
XX FT 197368..244442  
XX FT intron  
XX FT /tag= q  
XX FT /number= 8  
XX FT 244443..244695  
XX FT exon  
XX FT /tag= r  
XX FT /number= 9  
XX FT 244696..267106  
XX FT intron  
XX FT /tag= s  
XX FT /number= 9  
XX FT 267107..267303  
XX FT exon  
XX FT /tag= t  
XX FT /number= 10  
XX FT 267304..338564  
XX FT intron  
XX FT /tag= u  
XX FT /number= 10  
XX FT 338565..339509  
XX FT exon  
XX FT /tag= v  
XX FT /number= 11  
XX FT replace (2557,T)  
XX FT /tag= w  
XX FT variation

```
FT variation /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace (5492,C)
FT *tag= x
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace (5330,A)
FT *tag= y
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace (7269,C)
FT *tag= z
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace (9650,G/T)
FT *tag= aa
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace (10664,T)
FT *tag= ab
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace (11043,A/G/C)
FT *tag= ac
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace (13109,T)
FT *tag= ad
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace (13392,A)
FT *tag= ae
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace (13395..13397,GA)
FT *tag= af
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace (13563,C)
FT *tag= ag
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace (13694,A)
FT *tag= ah
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace (14075,A)
FT *tag= ai
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace (14573,A)
FT *tag= aj
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace (15013,T)
FT *tag= ak
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace (15073..15075,GT)
FT *tag= al
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace (15840,T)
FT *tag= am
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace (16346,T)
FT *tag= an
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace (16367,A)
FT *tag= ao
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace (16639,G)
FT *tag= ap
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace (18149,A/T)
FT *tag= aq
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace (18243,C/A)
FT *tag= ar
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace (18293,G/A)
FT *tag= as
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace (19185,C)
FT *tag= at
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace (24728..24730,AA)
FT *tag= au
FT /standard_name= "Single nucleotide polymorphism (SNP)"

FT variation replace (24790,T)
FT *tag= av
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace (24950,T)
FT *tag= aw
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace (28608,G)
FT *tag= ax
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace (29244,C)
FT *tag= ay
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace (29651,T)
FT *tag= az
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace (34034,T)
FT *tag= ba
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace (36671,G)
FT *tag= bb
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace (36836,T)
FT *tag= bc
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace (37010,A)
FT *tag= bd
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace (38461,A/C)
FT *tag= be
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace (38588,A/C)
FT *tag= bf
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace (39673,A)
FT *tag= bg
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace (39854,C)
FT *tag= bh
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace (41325,A)
FT *tag= bi
FT /standard_name= "Single nucleotide polymorphism (SNP)"
FT replace (41412,A)
FT *tag= bj
FT /standard_name= "Single nucleotide polymorphism (SNP)"

Query Match 23.5%; Score 28.2; DB 24; Length 341511;
Best Local Similarity 55.7%; Pred. No. 36;
Matches 54; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 21 AATCAAAATGCTGAGGACGAGCTGTGCTGCGAAGCAAACTCCTTACCGAGTCTTCT 80
DB 193970 AGTCATTGGCTGTTGGAGAGAGGGTCTATAGACTGAGTTCTTTTCGGATGCTTAG 194029

QY 81 ATACCCAGGCTCAATACCGCTCAAGAGAGAGCTAT 117
DB 194030 ATCCAAATAATGATTAGACATCCCAAGAGATAGGTAT 194066

Search completed: November 15, 2003, 00:35:24
JOB time : 81.4482 secs
```





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 14, 2003, 23:58:39 ; Search time 800.53 Seconds  
(without alignments)  
3643.257 Million cell updates/sec

Title: US-09-928-457-74

Perfect score: 120

Sequence: 1 CGGTCAAGACAGCGAAGCT.....CTCAAGGAGAGACTATCAT 120

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	31.8	26.5	545	9 AW669802	AW669802 113373 MA
C 2	31.2	25.0	589	12 BP009535	BP009535 BP009535
C 3	31.2	25.0	657	12 BP009682	BP009682 BP009682
4	30.4	25.3	479	29 FR002495	AL017805 F.rubripe

C 5	30.2	25.2	711	10 BF859869	BF859869 963014H01
C 6	30.2	25.2	736	10 BF859841	BF859841 963014G04
C 7	30.2	25.2	782	10 BF859840	BF859840 963014G04
C 8	30.2	25.2	797	10 BF859868	BF859868 963014H01
C 9	30	25.0	1035	29 BZ558760	BZ558760 963014H01
C 10	29.4	24.5	966	10 BF127560	BF127560 601810209
C 11	29.2	24.3	771	29 BZ966854	BZ966854 PUGK45TB
C 12	29.2	24.3	816	13 BU745275	BU745275 CHI008E
C 13	29	24.2	768	12 BI218807	BI218807 602938279
C 14	28.8	24.0	571	29 FR0026279	FR0026279
C 15	28.6	23.8	725	28 AQ960052	AQ960052 LERFB11TR
C 16	28.6	23.8	733	28 AQ960051	AQ960051 LERFB11TR
C 17	28.4	23.7	326	29 FR0024950	FR0024950
C 18	28.4	23.7	538	14 CB886149	CB886149 3529_1_2
C 19	28.4	23.7	601	28 BH814982	BH814982 AM_Ba001
C 20	28.4	23.7	665	29 AG127283	AG127283 Pan trogl
C 21	28.4	23.7	691	12 BU321620	BU321620 BU321620
C 22	28.4	23.7	693	12 BM865214	BM865214 mgap010XP
C 23	28.4	23.7	723	12 BU320632	BU320632 BU320632
C 24	28.2	23.5	597	9 AV593758	AV593758 AV593758
C 25	28	23.3	460	9 AA981455	AA981455 ua35f08.r
C 26	28	23.3	479	28 AZ048801	AZ048801 GSBBr0019
C 27	28	23.3	498	29 BZ603792	BZ603792 WHAAQ68TF
C 28	28	23.3	559	29 BZ606580	BZ606580 WHABB42TF
C 29	28	23.3	570	9 AL695123	AL695123 AL695123
C 30	28	23.3	676	28 BZ052416	BZ052416 jnr68f01
C 31	28	23.3	917	29 CC414833	CC414833 PUHPP45TB
C 32	28	23.3	980	29 CC213462	CC213462 CH261-115
C 33	28	23.3	1074	13 BU333735	BU333735 603869750
C 34	28	23.3	1278	13 BQ436845	BQ436845 AGENCOURT
C 35	27.8	23.2	234	29 FR0026265	FR0026265
C 36	27.8	23.2	360	13 BY071431	BY071431
C 37	27.8	23.2	400	10 BG101744	BG101744 EM1_5_G05
C 38	27.8	23.2	404	29 FR0023834	FR0023834
C 39	27.8	23.2	535	10 BG464307	BG464307 EM1_72_B0
C 40	27.8	23.2	553	9 AI866077	AI866077 wl25d07.x
C 41	27.8	23.2	594	29 FR0031004	FR0031004
C 42	27.8	23.2	707	12 BM683197	BM683197 UI-E-E01-
C 43	27.8	23.2	1181	13 BU253236	BU253236 603415556
C 44	27.6	23.0	472	29 CNS00070	CNS00070
C 45	27.6	23.0	482	14 H67236	H67236 yu65f05.r1

#### ALIGNMENTS

```

RESULT 1
AW669802 545 bp mRNA linear EST 25-APR-2001
LOCUS 113373 MARC 1BOV Bos taurus CDNA 5', mRNA sequence.
DEFINITION AW669802
ACCESSION AW669802
VERSION AW669802.1 GI:7526316
KEYWORDS EST.
SOURCE Bos taurus (cow)
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 545)
Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
Partea,G., Holt,I., Karanycheva,S., Liang,F., Quackenbush,J. and
Keele,J.W.
Sequence evaluation of four pooled-tissue normalized bovine CDNA
libraries and construction of a gene index for cattle
Genome Res. 11 (4), 626-630 (2001)
21180013
11282978
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366

```

TITLE  
JOURNAL  
MEDLINE  
PUBMED  
COMMENT

	Query Match	26.0%	Score 31.2	DB 12	Length 589
	Best Local Similarity	66.2%	Prod No. 11		
	Matches 45	Conservative 0	Mismatches 23	Indels 0	Gaps 0
27	ATGCTGAGCAGGACTGCTGCGAAGAAACTCTTACCGAAGTCCTATATCC	86			
94	AAAGCTTGAAGAACAAAGTGTTGTGAACCTAAATGCTTACCTATGGCAGCAATCG	35			
87	AGGCTCAA	94			
34	AGCTCAA	27			

```

BASE COUNT
130 a 142 c 116 g 157 t
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="MARC 1B0V"
/note="vector: pCMV SPORTS; Site 1: NotI; Site 2: SalI;
Library made from pooled tissue from lymph node, ovary,
fat, hypothalamus, and pituitary."

```

BP009682 657 bp mRNA linear EST 15-MAR-2002  
 BP009682 Nori Satoh unpublished cDNA library, young adult Ciona  
 intestinalis cDNA clone ctad7in03 5', mRNA sequence.  
 BP009682  
 BP009682.1 GI:19501159  
 EST.  
 Ciona intestinalis  
 Ciona intestinalis  
 Ciona intestinalis  
 Ciona intestinalis  
 Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;  
 Phlebobranchia; Cionidae; Ciona.

1 (bases 1 to 657)  
 Satoh,N., Satou,Y., Kohara,Y. and Shin-I,T.  
 Expressed genes in *Ciona intestinalis*  
 Unpublished  
 Contact: Nori Satoh  
 Department of Zoology  
 Kyoto University  
 Sakyo-ku, Kyoto, Kyoto 606-8502, Japan  
 Tel: 81-75-753-4081  
 Fax: 81-75-705-1113  
 E-mail: satoh@acidian.zool.kyoto-u.ac.jp.

```

FEATURES
    source
        Location/Qualifiers
            1..657
                /organism="Ciona intestinalis"
                /mol_type="mRNA"
                /db_xref="taxon:7719"
                /clone="ciad71n03"
                /tissue_type="whole animal"
                /dev_stage="young adult"
                /clone_lib="Nori Satoh unpublished cDNA library, young
                adult"

```

[illegible][illegible]

BASE COUNT  
ORIGIN

Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
Tetraodontidae; Tetraodontidae; Takifugu.  
1. (bases 1 to 479)  
Elgar, G., Clark, M., Smith, S., Meek, S., Warner, S., Umrani, Y.,  
Williams, G. and Brenner, S.  
Direct Submission  
Submitted (08-DEC-1997) MRC Human Genome Mapping Project Resource  
Centre Hinxton, Cambridge, CB10 1SB. Email: biohelp@hmp.mrc.ac.uk  
Vector: pBluescript II KS  
V-type: phagemid  
PRIMER: KS  
DESCR:  
One pass dye-terminator sequencing of cosmid cloned genomic  
sequence.

FEATURES  
source  
Location/Qualifiers  
1..479  
/organism="Takifugu rubripes"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:31033"  
/clone="135A08a6"  
/clone\_lib="cosmid 135A08"  
BASE COUNT 84 a 136 c 141 g 78 t 38 others  
ORIGIN

Query Match 25.3%; Score 30.4; DB 29; Length 479;  
Best Local Similarity 75.5%; Pred. No. 19;  
Matches 37; Conservative 0; Mismatches 12; Indels 0; Gaps 0;  
QY 61 ACTCCTACGAGTCTTCTATACCCAGGCTCAATAGCGCGTCAAGGAG 109  
Db 165 AGTCCATGAGAGCGCTTCTTACCAGGNTCAACTGCCTCTGAAGAG 213

RESULT 5  
BF859869/c  
LOCUS  
DEFINITION 711 bp mRNA linear EST 19-JAN-2001  
963014H01.y2 C. reinhardtii CC-1690, Stress condition I, normalized  
Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.  
ACCESSION  
BF859869  
VERSION  
BF859869.1 GI:12249962  
KEYWORDS  
EST.  
SOURCE  
Chlamydomonas reinhardtii  
Chlamydomonas reinhardtii  
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
Chlamydomonadaceae; Chlamydomonas.

REFERENCE  
AUTHORS Grossman, A., Davies, J., Federspiel, N., Harris, E., Hauser, C.,  
Lefebvre, P., McDermott, J.P., Shrager, J., Silflow, C. and Stern, D.  
TITLE Analyses of the Chlamydomonas reinhardtii Genome: A Model,  
Unicellular System for Analyzing Gene Function and Regulation in  
Vascular Plants; project phase 3  
JOURNAL  
COMMENT Unpublished  
Contact: Charles Hauser  
DCMB Box 91000  
Duke University  
Durham, NC 27708-1000  
Tel: 919 613 8159  
Fax: 919 613 8177  
Email: chauser@duke.edu.  
Location/Qualifiers  
1..711  
/organism="Chlamydomonas reinhardtii"  
/mol\_type="mRNA"  
/strain="CC-1690 wild type mt+ 21gr"  
/db\_xref="taxon:3055"  
/clone\_lib="C. reinhardtii CC-1690, Stress condition I,  
normalized, Lambda Zap II"  
/note="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2:  
XhoI. This library, constructed by John Davies and Jeffrey  
McDermott, combines cDNAs from CC-1690 cells grown to  
mid-log phase in TAP-N (30 min, 1hr, 4hr), TAP-S (30 min,  
1hr, 4hr), TAP-P (4hr, 12hr, 24hr), NO3 to NH4 (30min, 1hr  
4hr) and NH4 to NO3 (30min, 1hr, 4hr). PolyA mRNA was

FEATURES  
source  
Location/Qualifiers  
1..711  
/organism="Chlamydomonas reinhardtii"  
/mol\_type="mRNA"  
/strain="CC-1690 wild type mt+ 21gr"  
/db\_xref="taxon:3055"  
/clone\_lib="C. reinhardtii CC-1690, Stress condition I,  
normalized, Lambda Zap II"  
/note="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2:  
XhoI. This library, constructed by John Davies and Jeffrey  
McDermott, combines cDNAs from CC-1690 cells grown to  
mid-log phase in TAP-N (30 min, 1hr, 4hr), TAP-S (30 min,  
1hr, 4hr), TAP-P (4hr, 12hr, 24hr), NO3 to NH4 (30min, 1hr  
4hr) and NH4 to NO3 (30min, 1hr, 4hr). PolyA mRNA was

purified from each sample, pooled and cDNA synthesized.  
The cDNA was directionally cloned into lambda Zap II  
(Stratagene) in the EcoRI (5') and XhoRI (3') sites.  
pBluescript II SK- plasmids were excised from the lambda  
ZAP clones by superinfection with ExAssist (Stratagene)  
phage. The library was normalized using method 4 described  
in Bonaldo et al (1996) Genome Research 6: 791-806."

BASE COUNT 165 a 184 c 219 g 142 t 1 others  
ORIGIN

Query Match 25.2%; Score 30.2; DB 10; Length 711;  
Best Local Similarity 60.2%; Pred. No. 26;  
Matches 50; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 7 GAAACAGCAAGGTAATGAAAATCCTGAGCAGGACTGTGTCGGAACGAAACTCT 66  
Db 178 GAAACTAGCAGCAATGACAAAGGAAAGTGAGGACGACGTTGGGAAACAAAGCCTCG 119  
QY 67 TACCGAGTCTTCTATACCCAGG 89  
Db 118 TGCCGAATTCCTGAGCCCGGGG 96

RESULT 6  
BF859841/c  
LOCUS  
DEFINITION 736 bp mRNA linear EST 19-JAN-2001  
963014G04.y2 C. reinhardtii CC-1690, Stress condition I, normalized  
Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.  
ACCESSION  
BF859841  
VERSION  
BF859841.1 GI:12249934  
KEYWORDS  
EST.  
SOURCE  
Chlamydomonas reinhardtii  
Chlamydomonas reinhardtii  
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
Chlamydomonadaceae; Chlamydomonas.

REFERENCE  
AUTHORS Grossman, A., Davies, J., Federspiel, N., Harris, E., Hauser, C.,  
Lefebvre, P., McDermott, J.P., Shrager, J., Silflow, C. and Stern, D.  
TITLE Analyses of the Chlamydomonas reinhardtii Genome: A Model,  
Unicellular System for Analyzing Gene Function and Regulation in  
Vascular Plants; project phase 3  
JOURNAL  
COMMENT Unpublished  
Contact: Charles Hauser  
DCMB Box 91000  
Duke University  
Durham, NC 27708-1000  
Tel: 919 613 8159  
Fax: 919 613 8177  
Email: chauser@duke.edu.  
Location/Qualifiers  
1..736  
/organism="Chlamydomonas reinhardtii"  
/mol\_type="mRNA"  
/strain="CC-1690 wild type mt+ 21gr"  
/db\_xref="taxon:3055"  
/clone\_lib="C. reinhardtii CC-1690, Stress condition I,  
normalized, Lambda Zap II"  
/note="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2:  
XhoI. This library, constructed by John Davies and Jeffrey  
McDermott, combines cDNAs from CC-1690 cells grown to  
mid-log phase in TAP-N (30 min, 1hr, 4hr), TAP-S (30 min,  
1hr, 4hr), TAP-P (4hr, 12hr, 24hr), NO3 to NH4 (30min, 1hr  
4hr) and NH4 to NO3 (30min, 1hr, 4hr). PolyA mRNA was

FEATURES  
source  
Location/Qualifiers  
1..736  
/organism="Chlamydomonas reinhardtii"  
/mol\_type="mRNA"  
/strain="CC-1690 wild type mt+ 21gr"  
/db\_xref="taxon:3055"  
/clone\_lib="C. reinhardtii CC-1690, Stress condition I,  
normalized, Lambda Zap II"  
/note="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2:  
XhoI. This library, constructed by John Davies and Jeffrey  
McDermott, combines cDNAs from CC-1690 cells grown to  
mid-log phase in TAP-N (30 min, 1hr, 4hr), TAP-S (30 min,  
1hr, 4hr), TAP-P (4hr, 12hr, 24hr), NO3 to NH4 (30min, 1hr  
4hr) and NH4 to NO3 (30min, 1hr, 4hr). PolyA mRNA was

BASE COUNT 176 a 195 c 216 g 149 t  
ORIGIN

```

Query Match      25.2%; Score 30.2; DB 10; Length 736;
Best Local Similarity 60.2%; Pred. No. 27; Mismatches 0; Gaps 0;
Matches 50; Conservative 0; Indels 33;

QY      7  GAACAGCGCAGGTAATGAAATGCTCTGAGGCGACGCGACTGTCTCGAACGAAACTCT 66
      |||||
Db      180  GAACACTAGCGAGCAATGACAAGCGGAAAGTGAGGCAACGCTTGGGAAACAAGACCTCG 121

QY      67  TACCGAAGTCTTCTATACCCAGG 89
      |||||
Db      120  TGCCGAATTCCTGAGCCCGGG 98

RESULT 7
BF859840/c
LOCUS
DEFINITION
  BF859840.1 782 bp mRNA linear EST 19-JAN-2001
  Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION
  BF859840
VERSION
  BF859840.1 GI:12249933
KEYWORDS
  EST.
SOURCE
  Chlamydomonas reinhardtii
  Chlamydomonas reinhardtii
  Chlamydomonas reinhardtii
  Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
  Chlamydomonadales; Chlamydomonas.
REFERENCE
  1 (bases 1 to 782)
  Grossman A., Davies J., Federspiel N., Harris E., Hauser C.,
  Lefebvre P., McDermott J.P., Shrager J., Silflow C. and Stern D.
  Analyses of the Chlamydomonas reinhardtii Genome: A Model, in
  Cellular System for Analyzing Gene Function and Regulation in
  Vascular Plants; project phase 3
  Unpublished
  Contact: Charles Hauser
  DCMH Box 91000
  Duke University
  Durham, NC 27708-1000
  Tel: 919 613 8159
  Fax: 919 613 8177
  Email: chauser@duke.edu.

FEATURES
    source
        1..782
            /organism="Chlamydomonas reinhardtii"
            /mol_type="mRNA"
            /strain="CC-1690 wild type mt+ 21gr"
            /db_xref="taxon:3055"
            /clone_lib="C. reinhardtii CC-1690, Stress condition I,
            normalized, Lambda Zap II"
            /note="vector: pBluescript II SK-; Site 1: EcoRI; Site 2:
            XhoI; This library, constructed by John Davies and Jeffrey
            McDermott, combines cDNAs from CC-1690 cells grown to
            mid-log phase in TAP-N (30 min, 1hr, 4hr), TAP-S (30 min,
            1hr, 4hr), TAP-P (4hr, 12hr, 24hr), NO3 to NH4 (30min, 1hr,
            4hr) and NH4 to NO3 (30min, 1hr, 4hr). PolyA mRNA was
            purified from each sample, pooled and cDNA synthesized.
            The cDNA was directionally cloned into lambda Zap II
            (Stratagene) in the EcoRI (5') and XhoI (3') sites.
            pBluescript II SK- plasmids were excised from the lambda
            Zap clones by superinfection with EXAssist (Stratagene)
            phage. The library was normalized using method 4 described
            in Ronaldo et al (1996) Genome Research 6: 791-806."
        182 a 213 c 229 g 157 t 1 others
BASE COUNT
ORIGIN
    Query Match      25.2%; Score 30.2; DB 10; Length 782;
    Best Local Similarity 60.2%; Pred. No. 28;
    Matches 50; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

    QY      7  GAACAGCGCAGGTAATGAAATGCTCTGAGGCGCGACTGTCTCGAACGAAACTCT 66
    |||||
    Db      175  GAACACTAGCGAGCAATGACAAGCGGAAAGTGAGGCGACGCTTGGGAAACAAGACCTCG 116

    QY      67  TACCGAAGTCTTCTATACCCAGG 89
  
```

**Db** 115 TGCCGAATCTGCGACCGGGG 93  
**RESULT 8**  
**BP859868/c**  
**LOCUS** BP859868  
**DEFINITION** 963014H01.y1 C. reinhardtii CC-1690, Stress condition I, normalized  
, Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.  
**ACCESSION** BP859868  
**VERSION** BP859868.1 GI:12249961  
**KEYWORDS** EST.  
**SOURCE** Chlamydomonas reinhardtii  
**ORGANISM** Chlamydomonas reinhardtii  
Chlamydomonas reinhardtii  
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
Chlamydomonadales; Chlamydomonas.  
**REFERENCE** 1 (bases 1 to 797)  
**AUTHORS** Grossman, A., Davies, J., Federpiel, N., Harris, E., Hauser, C.,  
Lefebvre, P., McDermott, J.P., Shrager, J., Silflow, C. and Stern, D.  
**TITLE** Analysis of the Chlamydomonas reinhardtii Genome: A Model,  
Unicellular System for Analyzing Gene Function and Regulation in  
Vascular Plants; project phase 3  
**JOURNAL** Unpublished  
**COMMENT** Contact: Charles Hauser  
DCMB Box 91000  
Duke University  
Durham, NC 27708-1000  
Tel: 919 613 8159  
Fax: 919 613 8177  
Email: chauser@duke.edu.  
**FEATURES** Location/Qualifiers  
source 1..797  
/organism="Chlamydomonas reinhardtii"  
/mol\_type="mRNA"  
/strain="CC-1690 wild type wt+ 21gr"  
/db\_xref="taxon:3055"  
/clone\_lib="C. reinhardtii CC-1690, Stress condition I,  
normalized, Lambda Zap II"  
/note="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2: XhoI;  
This library, constructed by John Davies and Jeffrey  
McDermott, combines cDNAs from CC-1690 cells grown to  
mid-log phase in TAP-N (30 min, 1hr, 4hr), TAP-S (30 min,  
1hr, 4hr), TAP-P (4hr, 12hr, 24hr), NO3 to NH4 (30min, 1hr  
, 4hr) and NH4 to NO3 (30min, 1hr, 4hr). PolyA mRNA was  
purified from each sample, pooled and cDNA synthesized.  
The cDNA was directionally cloned into lambda Zap II  
(Stratagene) in the EcoRI (5') and XhoI (3') sites.  
pBluescript II SK- plasmids were excised from the lambda  
Zap clones by superinfection with ExAssist (Stratagene)  
phase. The library was normalized using method 4 described  
in Borraldo et al (1996) Genome Research 6: 791-806."  
BASE COUNT 188 a 215 c 229 g 160 t 5 others  
ORIGIN  
Query Match 25.2%; Score 30.2; DB 10; Length 797;  
Best Local Similarity 60.2%; Pred. No. 28;  
Matches 50; Conservative 0; Mismatches 33; Indels 0; Gaps 0;  
QY 7 GAACACGGAAGTAAATGAAATCGCTGAGGCACGACTGTGTCGCAACGAAATCTCT 66  
DB 179 GAAACTAGCGAGCAATGCAAGCGCAAGTGGAGCAGCGTTGGGAAACAAAGCTCG 120  
QY 67 TACCGAAGTCTTCTATACCCAGG 89  
DB 119 TGCCGAATCTGCGACCGGGG 97  
**RESULT 9**  
**BZ558760/c**  
**LOCUS** BZ558760  
**DEFINITION** pa98401.461.s1 pacs2.164 Pseudomonas aeruginosa genomic clone  
pa98401.461, genomic survey sequence.  
**ACCESSION** BZ558760  
**VERSION** BZ558760.1 GI:27173873

```

KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Pseudomonas aeruginosa
Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
1 (bases 1 to 1035)
Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
Burns,J.L., Kaul,R. and Olsen,M.V.
Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
J. Bacteriol., (2002) In press
Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.

FEATURES
Location/Qualifiers
1..1035
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="2-164"
/db_xref="taxon:287"
/clone="pa98401.451"
/clone_lib="pac52-164"
/note="clinical isolate 2-164 Whole genomic shotgun
library."

BASE COUNT
168 a 337 c 333 g 196 t 1 others

Query Match 25.0%; Score 30; DB 29; Length 1035;
Best Local Similarity 57.4%; Pred. No. 36;
Matches 54; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 19 GTAATGAAATGCTGAGCAGCGACTGTGCTGCGAAGAAATCTCTACCGAAGCTT 78
Db 920 GAAGTGTGAATGCTTGATACAAAGGCCGCTGGTGAATGCTTGTCTCAAGGAAGT 861

QY 79 CTATACCCAGGCTCAATAGCCCTCAAGGAGAGA 112
Db 860 CGACGCCGCCGCTTGGCAGCGCTGCCGGGA 827

RESULT 10
BF127560
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BF127560 966 bp mRNA linear EST 24-OCT-2000
601810209F1 NIH_MGC_46 Homo sapiens CDNA clone IMAGE:4053027 5',
mRNA sequence.
BF127560
BF127560.1 GI:10966600
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 966)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: c9apbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCW891 row: b column: 04
High quality sequence stop: 393.
Location/Qualifiers

FEATURES
source

Pseudomonas aeruginosa
Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
1 (bases 1 to 1035)
Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
Burns,J.L., Kaul,R. and Olsen,M.V.
Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
J. Bacteriol., (2002) In press
Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.

FEATURES
Location/Qualifiers
1..1035
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="2-164"
/db_xref="taxon:287"
/clone="pa98401.451"
/clone_lib="pac52-164"
/note="clinical isolate 2-164 Whole genomic shotgun
library."

BASE COUNT
168 a 337 c 333 g 196 t 1 others

Query Match 25.0%; Score 30; DB 29; Length 1035;
Best Local Similarity 57.4%; Pred. No. 36;
Matches 54; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 19 GTAATGAAATGCTGAGCAGCGACTGTGCTGCGAAGAAATCTCTACCGAAGCTT 78
Db 920 GAAGTGTGAATGCTTGATACAAAGGCCGCTGGTGAATGCTTGTCTCAAGGAAGT 861

QY 79 CTATACCCAGGCTCAATAGCCCTCAAGGAGAGA 112
Db 860 CGACGCCGCCGCTTGGCAGCGCTGCCGGGA 827

RESULT 10
BF127560
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BF127560 966 bp mRNA linear EST 24-OCT-2000
601810209F1 NIH_MGC_46 Homo sapiens CDNA clone IMAGE:4053027 5',
mRNA sequence.
BF127560
BF127560.1 GI:10966600
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 966)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: c9apbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCW891 row: b column: 04
High quality sequence stop: 393.
Location/Qualifiers

FEATURES
source

Pseudomonas aeruginosa
Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
1 (bases 1 to 1035)
Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
Burns,J.L., Kaul,R. and Olsen,M.V.
Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
J. Bacteriol., (2002) In press
Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.

FEATURES
Location/Qualifiers
1..1035
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="2-164"
/db_xref="taxon:287"
/clone="pa98401.451"
/clone_lib="pac52-164"
/note="clinical isolate 2-164 Whole genomic shotgun
library."

BASE COUNT
168 a 337 c 333 g 196 t 1 others

Query Match 25.0%; Score 30; DB 29; Length 1035;
Best Local Similarity 57.4%; Pred. No. 36;
Matches 54; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 19 GTAATGAAATGCTGAGCAGCGACTGTGCTGCGAAGAAATCTCTACCGAAGCTT 78
Db 920 GAAGTGTGAATGCTTGATACAAAGGCCGCTGGTGAATGCTTGTCTCAAGGAAGT 861

QY 79 CTATACCCAGGCTCAATAGCCCTCAAGGAGAGA 112
Db 860 CGACGCCGCCGCTTGGCAGCGCTGCCGGGA 827

RESULT 10
BF127560
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BF127560 966 bp mRNA linear EST 24-OCT-2000
601810209F1 NIH_MGC_46 Homo sapiens CDNA clone IMAGE:4053027 5',
mRNA sequence.
BF127560
BF127560.1 GI:10966600
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 966)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: c9apbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCW891 row: b column: 04
High quality sequence stop: 393.
Location/Qualifiers

FEATURES
source

Pseudomonas aeruginosa
Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
1 (bases 1 to 1035)
Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
Burns,J.L., Kaul,R. and Olsen,M.V.
Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
J. Bacteriol., (2002) In press
Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.

FEATURES
Location/Qualifiers
1..1035
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="2-164"
/db_xref="taxon:287"
/clone="pa98401.451"
/clone_lib="pac52-164"
/note="clinical isolate 2-164 Whole genomic shotgun
library."

BASE COUNT
168 a 337 c 333 g 196 t 1 others

Query Match 25.0%; Score 30; DB 29; Length 1035;
Best Local Similarity 57.4%; Pred. No. 36;
Matches 54; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 19 GTAATGAAATGCTGAGCAGCGACTGTGCTGCGAAGAAATCTCTACCGAAGCTT 78
Db 920 GAAGTGTGAATGCTTGATACAAAGGCCGCTGGTGAATGCTTGTCTCAAGGAAGT 861

QY 79 CTATACCCAGGCTCAATAGCCCTCAAGGAGAGA 112
Db 860 CGACGCCGCCGCTTGGCAGCGCTGCCGGGA 827

RESULT 10
BF127560
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BF127560 966 bp mRNA linear EST 24-OCT-2000
601810209F1 NIH_MGC_46 Homo sapiens CDNA clone IMAGE:4053027 5',
mRNA sequence.
BF127560
BF127560.1 GI:10966600
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 966)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: c9apbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCW891 row: b column: 04
High quality sequence stop: 393.
Location/Qualifiers

FEATURES
source

Pseudomonas aeruginosa
Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
1 (bases 1 to 1035)
Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
Burns,J.L., Kaul,R. and Olsen,M.V.
Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
J. Bacteriol., (2002) In press
Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.

FEATURES
Location/Qualifiers
1..1035
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="2-164"
/db_xref="taxon:287"
/clone="pa98401.451"

```

Query Match 24.3%; Score 29.2; DB 29; Length 771;  
 Best Local Similarity 57.8%; Pred. No. 59;  
 Matches 52; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 6 AGAAACAGCAGGAGTAATGAAATGCTGAGGACGAGTGTGCTGCGAAGCAAACTCC 65  
 DB 752 AAAAAGTGGCTGAGAAACAAACACCATGGGCAACATGAGATGAGCAGCTAATCCA 693

QY 66 TTACCGAAGTCTTCTATACCCAGGCTCAAT 95  
 DB 692 GTACCCAAAGTTTCAATACTAGAGCAAT 663

RESULT 12  
 LOCUS BU745275/c  
 DEFINITION CH1#008\_E04T7 Canine heart non-normalized cDNA Library in pBluescript Canis familiaris cDNA clone CH1#008\_E04 5', mRNA sequence.  
 ACCESSION BU745275  
 VERSION BU745275  
 KEYWORDS EST:  
 SOURCE Canis familiaris (dog)  
 ORGANISM Canis familiaris  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
 Y.Y., Desai, R., Olarte, M., Henthorn, P. and George A.L.  
 1 (bases 1 to 816)  
 TITLE Expressed sequence tags from Canine heart  
 JOURNAL Unpublished  
 COMMENT Other\_ESTs: CH1#008\_E04T3  
 Contact: George AL  
 Division of Genetic Medicine  
 Vanderbilt University  
 529 Light Hall, 2215 Garland Avenue, Nashville, TN 37232-0275, USA  
 Tel: 615 936 2660  
 Fax: 615 936 2661  
 Email: al.george@vanderbilt.edu  
 Insert Length: 1116 Std Error: 0.00  
 Seq primer: T7: TAATACGACTCACTATAGGG  
 High quality sequence start: 93  
 High quality sequence stop: 419.  
 Location/Qualifiers  
 1..816  
 /organism="Canis familiaris"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9615"  
 /clone="CH1#008\_E04"  
 /tissue\_type="heart"  
 /cell\_type="heart"  
 /dev\_stage="mixed developmental stages (adult, 30 day - 40 day fetal)"  
 /clone\_lib="Canine heart non-normalized cDNA Library in pBluescript"  
 /notes="Organ: heart; Vector: pBluescript; Site: 1: 5' of vector NotI; Site: 2: 3' of vector EcoRI; Tissue source: dog heart (adult, 30 day - 40 day fetal) right and left atria and ventricle. Dog breed - mixed (beagle, German shepherd, pointer, Irish setter). Library construction: oligo-dT primed"

BASE COUNT 171 a 250 c 214 g 180 t  
 ORIGIN

Query Match 24.3%; Score 29.2; DB 13; Length 816;  
 Best Local Similarity 59.8%; Pred. No. 61;  
 Matches 49; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 25 AAAATGCTGAGCAGGAGTGTGCTGCGAAGCAAACTCTTACCGAAGTCTTCTATAC 84  
 DB 265 AAAATCCCGAGCTCAGTCCGCGAGCGGAAACACCTACCGAACTTTCTTGAT 206

QY 85 CCAGGCTCAATAGCCGCTCAG 106

Db 205 GGAGCCTCGGAGCGGAAAAAG 184

RESULT 13  
 LOCUS BI218807  
 DEFINITION 602938279F1 NCI\_CGAP\_Li9 Mus musculus cDNA clone IMAGE:5101627 5', mRNA sequence.  
 ACCESSION BI218807  
 VERSION BI218807.1 GI:14572251  
 KEYWORDS EST:  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 768)  
 TITLE NIH-MGC http://mgs.nci.nih.gov/  
 JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
 COMMENT Unpublished  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgabs@mail.nih.gov  
 Tissue procurement: Jeffrey E. Green, M.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: L1AM11244 row: m column: 20  
 High quality sequence stop: 766.  
 Location/Qualifiers  
 1..768  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="FVB/N"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:5101627"  
 /lab\_host="NCI\_CGAP\_Li9"  
 /notes="Organ: liver; Vector: pCMV-SPORT6; Site: 1: NotI; Site: 2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.9 kb. Constructed by Life Technologies. Note: this is a NCI\_CGAP Library."  
 BASE COUNT 163 a 202 c 215 g 188 t  
 ORIGIN

Query Match 24.2%; Score 29; DB 12; Length 768;  
 Best Local Similarity 57.0%; Pred. No. 69;  
 Matches 53; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 13 GGCAAGCTAATGAAATGCTGAGGACGAGTGTGCTGCGAAGCAAACTCTTACCGA 72  
 DB 493 GGCTGGCGAAGAAATCACTGGAGACTGGGTGACAGGAGCCGACACCTCTGTGA 552

QY 73 AGTCTTATACCCAGGCTCAATAGCGCTCAA 105  
 DB 553 TGCCTTACCGCCGAGGCTGGCAGGCCATCAA 585

RESULT 14  
 LOCUS FR0026279  
 DEFINITION F.rubripes GSS sequence, clone 169K01ac9, genomic survey sequence.  
 ACCESSION AL019112  
 VERSION AL019112.1 GI:2685480  
 KEYWORDS GSS: genome survey sequence.  
 SOURCE Takifugu rubripes  
 ORGANISM Takifugu rubripes  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodontidae; Takifugu.  
 1 (bases 1 to 571)

AUTHORS Elgar,G., Clark,M., Smith,S., Meek,S., Warner,S., Umrانيا,Y.,  
Williams,G. and Brenner,S.  
TITLE Direct Submission  
JOURNAL Submitted (08-DEC-1997) MRC Human Genome Mapping Project Resource  
Centre Hinxton, Cambridge, CB10 1SB. Email: biohelp@gmp.mrc.ac.uk  
COMMENT Vector: pBluescript II KS  
V type: phagemid  
PRIMER: KS  
DESCR: One pass dye-terminator sequencing of cosmid cloned genomic  
sequence.

FEATURES Location/Qualifiers  
1..571  
/organism="Takifugu rubripes"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:31033"  
/clone="169K01aC9"  
/clone\_lib="cosmid 169K01"  
BASE COUNT 97 a 166 c 159 g 92 t 58 others  
ORIGIN

Query Match 24.0%; Score 28.8; DB 29; Length 571;  
Best Local Similarity 72.0%; Pred. No. 71;  
Matches 36; Conservative 0; Mismatches 14; Indels 0; Gaps 0;  
QY 59 AAATCCTTACCGAGTCTTCTATACCCAGGCTCAATAGCGCTCAAGGA 108  
Db 39 ANAGTCCATGAGAGGCTTCTTACCCAGGCTCAGCTCTCTGAAGA 88

RESULT 15  
AQ960052/c  
LOCUS 725 bp DNA linear GSS 28-JAN-2000  
DEFINITION LERPB11TR LERA Arabidopsis thaliana genomic clone LERPB11, genomic  
survey sequence.  
ACCESSION AQ960052  
VERSION AQ960052.1 GI:6787876  
KEYWORDS GSS.  
SOURCE Arabidopsis thaliana (thale cress)  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids  
; eurosids II; Brassicales; Brassicaceae; Arabidopsi.  
1 (bases 1 to 725)  
Buell,C.R., Lin,X., Pai,G., Barnstead,M., Bowman,C., Utterbach,T.,  
Feldblyum,T., Liang,F., Creasy,T. and Fraser,C.M.  
Genomic survey sequencing of Landsberg erecta ecotype of  
Arabidopsis thaliana and identification of sequence-based  
polymorphisms  
Unpublished  
Contact: Xiaoying Lin  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: at@tigr.org  
For additional information, see <http://www.tigr.org/tdb/at/at.html>  
Seq primer: TR  
Class: shotgun.

FEATURES Location/Qualifiers  
1..725  
/organism="Arabidopsis thaliana"  
/mol\_type="genomic DNA"  
/strain="Landsberg erecta"  
/db\_xref="taxon:3702"  
/clone="LERPB11"  
/clone\_lib="LERA"  
/notes="Organ: Leaf; Vector: pHS01; Total genomic DNA was  
sheared to 0.9-1 Kbp before ligation."  
BASE COUNT 236 a 133 c 169 g 187 t  
ORIGIN

Query Match 23.8%; Score 28.6; DB 28; Length 725;

Best Local Similarity 61.3%; Pred. No. 92;  
Matches 46; Conservative 0; Mismatches 29; Indels 0; Gaps 0;  
QY 13 GGCAAGGTAATGAAAAATGCCTGAGGCACGACTGTGCGAAACGAAAACTCCTTACCGA 72  
Db 602 GGCAGGTTAGTGTAATTTCTAACCTGTTGAGTAATCTTGAAGCAATCTCTCCAA 543  
QY 73 AGTCTTCTATACCCA 87  
Db 542 AATCTTCTGTAACA 528

Search completed: November 15, 2003, 08:02:24  
Job time : 803.53 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 15, 2003, 00:06:39 ; Search time 18.691 Seconds  
(without alignments)  
2833.774 Million cell updates/sec

Title: US-09-928-457-74

Perfect score: 120

Sequence: 1 CGGTAGACAGGCAAGGT.....CTCAAGGAGAGACTATCAT 120

Scoring table: IDENTITY NUC

Gapop 10.0 , Capext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:\*

- 1: /cgn2\_6/prodata/2/ina/5A-COMB.seq:\*
- 2: /cgn2\_6/prodata/2/ina/5B-COMB.seq:\*
- 3: /cgn2\_6/prodata/2/ina/6A-COMB.seq:\*
- 4: /cgn2\_6/prodata/2/ina/6B-COMB.seq:\*
- 5: /cgn2\_6/prodata/2/ina/PTUS-COMB.seq:\*
- 6: /cgn2\_6/prodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	27.2	22.7	24317	2	US-08-846-762-1
2	26.4	22.0	1821	1	US-08-229-287-3
3	26.4	22.0	8543	2	US-08-496-944-1
4	26.2	21.8	5543	2	US-08-687-080-101
5	25.4	21.2	657	3	US-09-385-982-326
6	25.4	21.2	5303	4	US-08-971-395-4
7	25.4	21.2	9168	4	US-09-687-731-11
8	25	20.8	2266	1	US-08-453-472-1
9	25	20.8	2266	1	US-08-453-952-1
10	25	20.8	2266	2	US-08-484-993B-42
11	25	20.8	2266	2	US-08-862-903-1
12	25	20.8	2266	2	US-08-484-158B-42
13	25	20.8	2266	2	US-08-484-596A-42
14	25	20.8	2266	2	US-08-480-150A-42
15	25	20.8	2266	3	US-08-458-731-42
16	25	20.8	2266	3	US-08-149-223A-42
17	24.8	20.7	1230025	4	US-09-199-452A-1
18	24.6	20.5	1266	4	US-09-252-991A-1214
19	24.6	20.5	2262	4	US-09-252-991A-1353
20	24.6	20.5	2418	4	US-09-252-991A-1251
21	24.6	20.5	3910	4	US-09-120-653D-1
22	24.6	20.5	4403765	3	US-08-103-840A-2
23	24.6	20.5	4411529	3	US-09-103-840A-1
24	24.4	20.3	900	4	US-09-589-927-1
25	24.4	20.3	900	4	US-09-277-665-1
26	24.4	20.3	900	4	US-09-589-987-1
27	24.4	20.3	1077	4	US-09-252-991A-12157

C	28	24.4	20.3	1835	3	US-09-276-531-26	Sequence 26, Appl
	29	24.4	20.3	4103	4	US-09-620-312D-390	Sequence 390, App
	30	24.4	20.3	6674	4	US-09-620-312D-110	Sequence 110, App
C	31	24.2	20.2	715	2	US-08-211-718-1	Sequence 1, Appli
C	32	24.2	20.2	1609	2	US-08-211-718-6	Sequence 6, Appli
C	33	24.2	20.2	2257	1	US-08-452-722-6	Sequence 6, Appli
C	34	24.2	20.2	2257	1	US-08-404-731A-6	Sequence 6, Appli
C	35	24.2	20.2	2257	1	US-08-344-227-6	Sequence 6, Appli
C	36	24.2	20.2	2257	2	US-08-503-226B-6	Sequence 6, Appli
C	37	24.2	20.2	2257	3	US-08-721-458B-6	Sequence 6, Appli
C	38	24	20.0	2115	1	US-08-285-641-19	Sequence 19, Appl
C	39	24	20.0	2418	4	US-09-388-743-25	Sequence 25, Appl
C	40	23.6	19.7	1602	4	US-09-107-532A-1590	Sequence 1590, Ap
C	41	23.6	19.7	1858	3	US-08-742-185-96	Sequence 96, Appl
C	42	23.6	19.7	2169	1	US-08-379-496-1	Sequence 1, Appli
C	43	23.6	19.7	15239	1	US-08-390-878-17	Sequence 17, Appl
C	44	23.6	19.7	21119	4	US-09-453-702B-111	Sequence 111, App
C	45	23.6	19.7	4403765	3	US-09-103-840A-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1  
US-08-846-762-1  
; Sequence 1, Application US/08846762A  
; Patent No. 5994072  
; GENERAL INFORMATION:  
; APPLICANT: Lam, Joseph S.  
; APPLICANT: Burrows, Lori  
; APPLICANT: Charter, Deborah  
; APPLICANT: de Kievit, Teresa  
; TITLE OF INVENTION: No. 5994072el Proteins Involved in the Synthesis and Assembly  
; TITLE OF INVENTION: of O-Antigen in Pseudomonas Aeruginosa  
; FILE REFERENCE: 6580-089  
; CURRENT APPLICATION NUMBER: US/08/846.762A  
; CURRENT FILING DATE: 1997-04-30  
; NUMBER OF SEQ ID NOS: 100  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 24417  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-08-846-762-1

Query Match	22.7%	Score 27.2;	DB 2;	Length 24417;
Best Local Similarity	61.1%	Pred. No. 3.1;		
Matches	44;	Conservative	0;	Mismatches 28; Indels 0; Gaps 0;
Qy	37	GCACGGACTGCTGCGACGAGAAAACCTCTTACCGAAGTCTTCTATACCCAGGCTCAATA	96	
Db	14511	GCACGGACTAGCCTGGGAAGACCATTCAGACGATGCTCAGCTTTACGGAGCGGTCAA	14570	

Qy	97	GCCTCTCAAGGA	108
Db	14571	GCCTCTCTCGGA	14582

RESULT 2  
US-08-229-287-3  
; Sequence 3, Application US/08229287  
; Patent No. 5530193  
; GENERAL INFORMATION:  
; APPLICANT: Clark Jr., John M.  
; APPLICANT: Jilka, Joseph M.  
; APPLICANT: Murry, Lynn E.  
; APPLICANT: Scarafia, Liliana E.  
; TITLE OF INVENTION: VIRUS RESISTANT CORN PLANTS  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sandoz Agro, Inc.  
; STREET: 975 California Avenue  
; CITY: Palo Alto



```
; GENERAL INFORMATION:
; APPLICANT: Gregory Dolgiov
; TITLE OF INVENTION: Human RAD50 Gene and Methods of Use Thereof
; NUMBER OF SEQUENCES: 175
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/687,080
; FILING DATE: 17-JUL-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/592,126
; FILING DATE: 26-JAN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Sholtz, Charles K.
; REGISTRATION NUMBER: 38,615
; REFERENCE/DOCKET NUMBER: 4600-0111.30
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 101:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5543 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEetical: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: 3' END OF INTRON 19 OF RAD50 GENOMIC
; INDIVIDUAL ISOLATE: SEQUENCE
; US-08-687-080-101

Query Match 21.8%; Score 26.2; DB 2; Length 5543;
Best Local Similarity 60.8%; Pred. No. 4;
Matches 43; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 9 AACAGGCAAGGTAATGAAATGCTGAGGCAGGACTGTGCTGCGAAGCAAACTCCTTA 68
|||
Db 3056 AAAAGGCATTTATTAGCAATTATTGAGTCAGGCACCTATCTGATTACATACATTC 3115
|||
QY 69 CCGAAGCTCTTC 79
|||
Db 3116 CTTTAATCTTC 3126
|||

RESULT 5
US-09-385-982-326
; Sequence 326, Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCDNA-250XX
; CURRENT APPLICATION NUMBER: US/09/385,982
; CURRENT FILING DATE: 1999-08-30
; EARLIER APPLICATION NUMBER: 09/328,111
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: 60/117,393
; EARLIER FILING DATE: 1999-01-27
; EARLIER APPLICATION NUMBER: 60/098,639
```

```
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 544
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 326
; LENGTH: 657
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(657)
; OTHER INFORMATION: n = A,T,C or G
; US-09-385-982-326

Query Match 21.2%; Score 25.4; DB 3; Length 657;
Best Local Similarity 64.8%; Pred. No. 3.3;
Matches 35; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 41 GGACTGTCTGCGAAGCAAACTCCTTACCGAAGTCTTCTATACCCAGGCTCAA 94
|||
Db 335 GGAATGCAATTGGCATGAACNCAGCTTACTGNAGNCTTCCAAACCTGGGCTCAA 388
|||

RESULT 6
US-08-971-395-4
; Sequence 4, Application US/08971395
; Patent No. 6359197
; GENERAL INFORMATION:
; APPLICANT: Amasino, Richard M
; APPLICANT: No. 6359197, Yoo-Sun
; APPLICANT: Gan, Susheng
; TITLE OF INVENTION: Transgenic Plants with Altered
; TITLE OF INVENTION: Senescence Characteristics
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/971,395
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J
; REGISTRATION NUMBER: 27386
; REFERENCE/DOCKET NUMBER: 960296.94908
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 608-251-5000
; TELEFAX: 608-251-9166
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5303 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-971-395-4

Query Match 21.2%; Score 25.4; DB 4; Length 5303;
Best Local Similarity 54.9%; Pred. No. 7.9;
Matches 50; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 16 AAGTAAATGAAATGCTGAGGCAGGACTGTGCTGCGAAGCAAACTCCTTACCGAAGT 75
|||
Db 4890 AAAGATATCAAGGTAAGAGGATTATGTGCTTCCATGAAGGCTTCTTACCCCACT 4949
|||
```

```
QY 76 CTTCTATACCCAGGCTCAATAGCGCTCAAG 106
Db 4950 ATATGAAACCCGGTTCAATACCCGGTCAAG 4980

RESULT 7
US-09-687-731-11
; Sequence 11, Application US/09687731
; Patent No. 6504080
; GENERAL INFORMATION:
; APPLICANT: Van Der Putten, Petrus Herman Maria
; TITLE OF INVENTION: Transgenic animal model for neurodegenerative disorders
; FILE REFERENCE: 'alpha-synuclein transgenic animals
; CURRENT APPLICATION NUMBER: US/09/687,731
; CURRENT FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: GB 9924513.6
; PRIOR FILING DATE: 1999-10-15
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 9168
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:DNA-construct
US-09-687-731-11

Query Match 21.2%; Score 25.4; DB 4; Length 9168;
Best Local Similarity 53.5%; Pred. No. 10; Mismatches 46; Indels 0; Gaps 0;
Matches 53; Conservative 0;

QY 5 CAGAACAGCGAAGGTATGAAATGCCCTGAGGCACGGACTGTGTCGGAACGAAACTC 64
Db 6447 CAGTAGCATGGAACATGATGAAACCAACCAATCCATTCCCTTTGCTGATATAACAGGCTC 6506

QY 65 CTTACCGAAGTCTTCTATACCCAGGCTCAATAGCGCTC 103
Db 6507 CAAAGCAAAACCTGTCTACTGAGGCTCAAGAGCAGATC 6545

RESULT 8
US-08-453-472-1
; Sequence 1, Application US/08453472
; Patent No. 5626846
; GENERAL INFORMATION:
; APPLICANT: DEAN, JURRIEN
; TITLE OF INVENTION: CONTRACEPTIVE VACCINE
; FILE REFERENCE: BASED ON ALLOIMMUNIZATION WITH ZONA PELLUCIDA
; TITLE OF INVENTION: POLYPEPTIDES
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/453,472
; FILING DATE: 30-May-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/038,948
; FILING DATE: 26-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/930,462
; FILING DATE: 20-AUG-1992
; PRIOR APPLICATION DATA:
```

```
; APPLICATION NUMBER: US 07/364,379
; FILING DATE: 12-JUN-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: DOROTHY R. AUTH
; REGISTRATION NUMBER: 36,434
; REFERENCE/DOCKET NUMBER: 2026-4032 US3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2266
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: human
; STRAIN:
; INDIVIDUAL ISOLATE:
; DEVELOPMENTAL STAGE:
; HAPLOTYPE:
; TISSUE TYPE:
; CELL TYPE:
; CELL LINE:
; ORGANELLE:
; FEATURE:
; NAME/KEY: ZP2
; LOCATION:
; IDENTIFICATION METHOD:
; OTHER INFORMATION: human ZP2 cDNA
US-08-453-472-1

Query Match 20.8%; Score 25; DB 1; Length 2266;
Best Local Similarity 51.3%; Pred. No. 7.9;
Matches 58; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 6 AGAACAGCGAAGGTATGAAATGCCCTGAGGCACGGACTGTGTCGGAACGAAACTCC 65
Db 918 AGAAGCAACAAATGGCATGAAATTCATTTCAAAAACTCTGCTCAAAACGAAATTATC 977

QY 66 TTACCGAAGTCTTCTATACCCAGGCTCAATAGCGCTCAAGGAGAGAGCTATC 118
Db 978 TGAAAAATGCCTACTCCATCAGTTCTACTTAGCTTCATCAAGCTGACCTTTC 1030

RESULT 9
US-08-453-952-1
; Sequence 1, Application US/08453952
; Patent No. 5672488
; GENERAL INFORMATION:
; APPLICANT: DEAN, JURRIEN
; TITLE OF INVENTION: CONTRACEPTIVE VACCINE
; FILE REFERENCE: BASED ON ALLOIMMUNIZATION WITH ZONA PELLUCIDA
; TITLE OF INVENTION: POLYPEPTIDES
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/453,952
; FILING DATE: 30-May-1995
; CLASSIFICATION: 424
```

```

/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 08/038,948
/ FILING DATE: 26-MAR-1993
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 07/930,462
/ FILING DATE: 20-AUG-1992
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 07/364,379
/ FILING DATE: 12-JUN-1989
/ ATTORNEY/AGENT INFORMATION:
/ NAME: DOROTHY R. AUTH
/ REGISTRATION NUMBER: 36,434
/ REFERENCE/DOCKET NUMBER: 2026-4032 US4
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (212) 758-4300
/ TELEFAX: (212) 751-6849
/ TELEX: 421792
/ INFORMATION FOR SEQ ID NO: 1:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 2266
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: unknown
/ MOLECULE TYPE: cDNA
/ ORIGINAL SOURCE:
/ ORGANISM: human
/ STRAIN:
/ INDIVIDUAL ISOLATE:
/ DEVELOPMENTAL STAGE:
/ HAPLOTYPE:
/ TISSUE TYPE:
/ CELL TYPE:
/ CELL LINE:
/ ORGANELLE:
/ FEATURE:
/ NAME/KEY: ZP2
/ LOCATION:
/ IDENTIFICATION METHOD:
/ OTHER INFORMATION: human ZP2 cDNA
US-08-453-952-1

```

```

Query Match      20.8%; Score 25; DB 1; Length 2266;
Best Local Similarity 51.3%; Pred. No. 7.9;
Matches 58; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY      6 AGAAGCAGGCAAGTAATGAAATGCTGAGGCAAGGACTGTGCTGCGAAGCAAACTCC 65
Db      918 AGAAGCAACAATGGCATGAAATGCAATTCAGCAAACTCTGCTCAAAACGAAATTTATC 977

QY      66 TTACCGAAGTCTTCTATACCCAGGCTCAATAGCCGCTCAAGGAGAGAGCTATC 118
Db      978 TGAAAAATGCTACTCCATCAGTTCTACTTAGCTTCACTCAAGCTGACCTTTC 1030

```

```

RESULT 10
US-08-484-993B-42
/ Sequence 42, Application US/08484993B
/ Patent No. 5837497
/ GENERAL INFORMATION:
/ APPLICANT: Harris Ph.D., Jeffrey D.
/ APPLICANT: Hsu, Kuang T.
/ APPLICANT: Podolski, Joseph S.
/ TITLE OF INVENTION: Materials and Methods for Immunococontraception
/ NUMBER OF SEQUENCES: 59
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
/ STREET: 6300 Sears Tower, 233 South Wacker Drive
/ CITY: Chicago
/ STATE: Illinois
/ COUNTRY: United States of America
/ ZIP: 60606-6402
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk

```

```

/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent In Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/484,993B
/ FILING DATE: 09-NOV-1993
/ CLASSIFICATION: 424
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 08/012,990
/ FILING DATE: 29-JAN-1993
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 07/973,341
/ FILING DATE: 09-NOV-1992
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Clough, David W.
/ REGISTRATION NUMBER: 36,107
/ REFERENCE/DOCKET NUMBER: 31745
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 312/474-6653
/ TELEFAX: 312/474-0448
/ TELEX: 25-3856
/ INFORMATION FOR SEQ ID NO: 42:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 2266 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: cDNA
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 1..2235
US-08-484-993B-42

```

```

Query Match      20.8%; Score 25; DB 2; Length 2266;
Best Local Similarity 51.3%; Pred. No. 7.9;
Matches 58; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY      6 AGAAGCAGGCAAGTAATGAAATGCTGAGGCAAGGACTGTGCTGCGAAGCAAACTCC 65
Db      918 AGAAGCAACAATGGCATGAAATGCAATTCAGCAAACTCTGCTCAAAACGAAATTTATC 977

QY      66 TTACCGAAGTCTTCTATACCCAGGCTCAATAGCCGCTCAAGGAGAGAGCTATC 118
Db      978 TGAAAAATGCTACTCCATCAGTTCTACTTAGCTTCACTCAAGCTGACCTTTC 1030

```

```

RESULT 11
US-08-862-903-1
/ Sequence 1, Application US/08862903
/ Patent No. 5916768
/ GENERAL INFORMATION:
/ APPLICANT: DEAN, JURRIEN
/ TITLE OF INVENTION: CONTRACEPTIVE VACCINE
/ TITLE OF INVENTION: BASED ON ALLOIMMUNIZATION WITH ZONA PELLUCIDA
/ NUMBER OF SEQUENCES: 12
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: MORGAN & FINNEGAN
/ STREET: 345 PARK AVENUE
/ CITY: NEW YORK
/ STATE: NEW YORK
/ COUNTRY: USA
/ ZIP: 10154
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: FLOPPY DISK
/ COMPUTER: IBM PC COMPATIBLE
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: WORDPERFECT 5.1
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/862,903
/ FILING DATE: 30-MAY-1995
/ CLASSIFICATION: 424
/ PRIOR APPLICATION DATA:

```

us-09-928-457-74.rni

Sun Nov 16 12:56:58 2003

```

APPLICATION NUMBER: US 08/038,948
FILING DATE: 26-MAR-1993
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/930,462
FILING DATE: 20-AUG-1992
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/364,379
FILING DATE: 12-JUN-1989
ATTORNEY/AGENT INFORMATION:
NAME: DOROTHY R. AUTH
REGISTRATION NUMBER: 36,434
REFERENCE/DOCKET NUMBER: 2026-4032 US4
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-6800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2266
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: cdna
ORGANISM: human
STRAIN:
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE:
HAPLOTYPE:
TISSUE TYPE:
CELL TYPE:
CELL LINE:
ORGANELLE:
FEATURE:
NAME/KEY: ZP2
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION: human ZP2 cdna
US-08-862-903-1

```

```

Query Match      20.8%; Score 25; DB 2; Length 2266;
Best Local Similarity 51.3%; Pred. No. 7.9;
Matches 58; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY      6 AGAACAGGCAAGTAAATGAAATGCTGAGGCACGACTGTGCTGCGAAGCAAACTCC 65
Db      918 AGAGCAACAAATGGCATGAATGCAATTCAGCAAACTCTGCTCAAAACGAATTATC 977

QY      66 TTACCGAAGTCTTCTATACCCAGGCTCAATAGCCGCTCAAGGAGAGAGCTATC 118
Db      978 TGAATAATGCTACTCCATCAGTTCTACTTAGCTTCACTCAAGCTGACCTTTC 1030

```

```

RESULT 12
US-08-484-158B-42
Sequence 42, Application US/08484158B
Patent No. 5976545
GENERAL INFORMATION:
APPLICANT: Harris Ph.D., Jeffrey D.
APPLICANT: Hsu, Kuang T.
APPLICANT: Podolski, Joseph S.
TITLE OF INVENTION: Pharmaceutical Compositions for
TITLE OF INVENTION: Immunoccontraception
NUMBER OF SEQUENCES: 61
CORRESPONDENCE ADDRESS:
ADDRESSEE: Borun
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:

```

```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,158B
FILING DATE: 07-JUNE-95
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/149,223
FILING DATE: 09-NOV-93
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/012,990
FILING DATE: 29-JAN-93
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/973,341
FILING DATE: 09-NOV-92
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.
REGISTRATION NUMBER: 36,107
REFERENCE/DOCKET NUMBER: 32794
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6653
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 2266 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cdna
FEATURE:
NAME/KEY: CDS
LOCATION: 1...2235
US-08-484-158B-42

```

```

Query Match      20.8%; Score 25; DB 2; Length 2266;
Best Local Similarity 51.3%; Pred. No. 7.9;
Matches 58; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY      6 AGAACAGGCAAGTAAATGAAATGCTGAGGCACGACTGTGCTGCGAAGCAAACTCC 65
Db      918 AGAGCAACAAATGGCATGAATGCAATTCAGCAAACTCTGCTCAAAACGAATTATC 977

QY      66 TTACCGAAGTCTTCTATACCCAGGCTCAATAGCCGCTCAAGGAGAGAGCTATC 118
Db      978 TGAATAATGCTACTCCATCAGTTCTACTTAGCTTCACTCAAGCTGACCTTTC 1030

```

```

RESULT 13
US-08-484-596A-42
Sequence 42, Application US/08484596A
Patent No. 5981228
GENERAL INFORMATION:
APPLICANT: Harris Ph.D., Jeffrey D.
APPLICANT: Hsu, Kuang T.
APPLICANT: Podolski, Joseph S.
TITLE OF INVENTION: Materials and Methods for Immunoccontraception
NUMBER OF SEQUENCES: 59
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

```

```

; APPLICATION NUMBER: US/08/484,596A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/149,223
; FILING DATE: 11-NOV-1993
; PRIOR APPLICATION DATA: 07/973,341
; APPLICATION NUMBER: 07/973,341
; FILING DATE: 09-NOV-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Clough, David W.
; REGISTRATION NUMBER: 36,107
; REFERENCE/DOCKET NUMBER: 31745
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6653
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2266 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..2235
; US-08-484-596A-42

Query Match 20.8%; Score 25; DB 2; Length 2266;
Best Local Similarity 51.3%; Pred. No. 7.9;
Matches 58; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 6 AGAAGCAGGCAAGTAAATGCAATGCTGAGGACGAGTGTGCTCGAAGCAAACTCC 65
Db 918 AGAAGCAAAATGGCATGAAATTCATTTGCAACAAACTCTGCTCAAAACGAAATTATC 977

QY 66 TTACCGAAGTCTTTATACCCAGGCTCAATAGCGCTCAAGGAGAGAGCTATC 118
Db 978 TGAATAATGCTACTCCATCAGTTCTACTTAGTCTCAAGCTGACCTTTC 1030

RESULT 14
US-08-480-150A-42
; Sequence 42, Application US/08480150A
; Patent No. 5989550
; GENERAL INFORMATION:
; APPLICANT: Harris Ph.D., Jeffrey D.
; APPLICANT: Hsu, Kuang T.
; APPLICANT: Podolski, Joseph S.
; TITLE OF INVENTION: Materials and Methods for Immunocontraception
; NUMBER OF SEQUENCES: 59
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,150A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/149,223
; FILING DATE: 09-NOV-1993
; APPLICATION NUMBER: 08/012,990
; FILING DATE: 29-JAN-1993

```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/973,341
; FILING DATE: 09-NOV-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Clough, David W.
; REGISTRATION NUMBER: 36,107
; REFERENCE/DOCKET NUMBER: 31745
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6653
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2266 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..2235
; US-08-480-150A-42

Query Match 20.8%; Score 25; DB 2; Length 2266;
Best Local Similarity 51.3%; Pred. No. 7.9;
Matches 58; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 6 AGAAGCAGGCAAGTAAATGCAATGCTGAGGACGAGTGTGCTCGAAGCAAACTCC 65
Db 918 AGAAGCAAAATGGCATGAAATTCATTTGCAACAAACTCTGCTCAAAACGAAATTATC 977

QY 66 TTACCGAAGTCTTTATACCCAGGCTCAATAGCGCTCAAGGAGAGAGCTATC 118
Db 978 TGAATAATGCTACTCCATCAGTTCTACTTAGTCTCAAGCTGACCTTTC 1030

RESULT 15
US-08-458-731-42
; Sequence 42, Application US/08458731
; Patent No. 6001599
; GENERAL INFORMATION:
; APPLICANT: Harris Ph.D., Jeffrey D.
; APPLICANT: Hsu, Kuang T.
; APPLICANT: Podolski, Joseph S.
; TITLE OF INVENTION: Materials and Methods for Immunocontraception
; NUMBER OF SEQUENCES: 59
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/458,731
; FILING DATE: 09-NOV-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/012,990
; FILING DATE: 29-JAN-1993
; APPLICATION NUMBER: 07/973,341
; FILING DATE: 09-NOV-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Clough, David W.
; REGISTRATION NUMBER: 36,107
; REFERENCE/DOCKET NUMBER: 31745
; TELECOMMUNICATION INFORMATION:

```

Search completed: November 15, 2003, 08:09:00  
Job time : 26.691 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 15, 2003, 00:09:34 ; Search time 67.9336 Seconds  
(without alignments)  
5771.357 Million cell updates/sec

Title: US-09-928-457-74

Perfect score: 120

Sequence: 1 CGGTCAGAACAGCGAAGT.....CTCAAGGAGAGAGCTATCAT 120

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 2169961 seqs, 1634102185 residues

Total number of hits satisfying chosen parameters: 4339922

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.\*

- 1: /cgn2\_6/ptodata1/pubna/US07\_PUBCOMB.seq\*
- 2: /cgn2\_6/ptodata1/pubna/PTCT\_NEW\_PUB.seq\*
- 3: /cgn2\_6/ptodata1/pubna/US06\_NEW\_PUB.seq\*
- 4: /cgn2\_6/ptodata1/pubna/US06\_PUBCOMB.seq\*
- 5: /cgn2\_6/ptodata1/pubna/US07\_NEW\_PUB.seq\*
- 6: /cgn2\_6/ptodata1/pubna/PTCT\_PUBCOMB.seq\*
- 7: /cgn2\_6/ptodata1/pubna/US08\_NEW\_PUB.seq\*
- 8: /cgn2\_6/ptodata1/pubna/US08\_PUBCOMB.seq\*
- 9: /cgn2\_6/ptodata1/pubna/US09A\_PUBCOMB.seq\*
- 10: /cgn2\_6/ptodata1/pubna/US09B\_PUBCOMB.seq\*
- 11: /cgn2\_6/ptodata1/pubna/US09C\_PUBCOMB.seq\*
- 12: /cgn2\_6/ptodata1/pubna/US09\_NEW\_PUB.seq\*
- 13: /cgn2\_6/ptodata1/pubna/US10A\_PUBCOMB.seq\*
- 14: /cgn2\_6/ptodata1/pubna/US10B\_PUBCOMB.seq\*
- 15: /cgn2\_6/ptodata1/pubna/US10\_NEW\_PUB.seq\*
- 16: /cgn2\_6/ptodata1/pubna/US60\_NEW\_PUB.seq\*
- 17: /cgn2\_6/ptodata1/pubna/US60\_PUBCOMB.seq\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	120	100.0	10	US-09-928-457-73
2	120	100.0	10	US-09-928-457-74
3	28	23.3	2598	US-09-938-842A-1646
4	27.6	23.0	261	US-09-923-876-5811
5	27.2	22.7	24417	US-10-216-209-1
6	27	22.5	525	US-10-027-632-51299
7	27	22.5	525	US-10-027-632-51299
8	26.6	22.2	394	US-09-960-352-13696
9	26.2	21.8	631	US-10-027-632-54625
10	26.2	21.8	631	US-10-027-632-54625
11	25.8	21.5	676	US-10-027-632-244583
12	25.8	21.5	676	US-10-027-632-244583
13	25.8	21.5	676	US-10-027-632-244583
14	25.8	21.5	676	US-10-027-632-244584
15	25.6	21.3	876	US-10-027-632-8784
16	25.6	21.3	876	US-10-027-632-8785

17	25.6	21.3	876	13	US-10-027-632-8784	Sequence 8784, Ap
18	25.6	21.3	876	13	US-10-027-632-8785	Sequence 8785, Ap
19	25.4	21.2	263	10	US-09-878-574-15498	Sequence 15498, A
20	25.4	21.2	565	12	US-10-027-632-234053	Sequence 234053, A
21	25.4	21.2	565	12	US-10-027-632-234054	Sequence 234054, A
22	25.4	21.2	565	13	US-10-027-632-234053	Sequence 234053, A
23	25.4	21.2	565	13	US-10-027-632-234054	Sequence 234054, A
24	25.4	21.2	565	11	US-09-871-161-326	Sequence 326, App
25	25.4	21.2	1027	11	US-09-822-846-69	Sequence 69, Appl
26	25.4	21.2	2121	12	US-10-027-632-250460	Sequence 250460, A
27	25.4	21.2	2121	12	US-10-027-632-250461	Sequence 250461, A
28	25.4	21.2	2121	13	US-10-027-632-250460	Sequence 250460, A
29	25.4	21.2	2121	13	US-10-027-632-250461	Sequence 250461, A
30	25.4	21.2	3378	10	US-09-917-800A-1443	Sequence 1443, Ap
31	25.2	21.0	565	12	US-10-029-386-10901	Sequence 10901, A
32	25.2	21.0	565	12	US-10-027-632-191607	Sequence 191607, A
33	25.2	21.0	565	12	US-10-027-632-191608	Sequence 191608, A
34	25.2	21.0	565	13	US-10-027-632-191607	Sequence 191607, A
35	25.2	21.0	565	13	US-10-027-632-191608	Sequence 191608, A
36	25.2	21.0	591	9	US-09-864-761-7709	Sequence 7709, Ap
37	25.2	21.0	987	14	US-10-156-761-198	Sequence 198, App
38	25.2	21.0	9025608	14	US-10-156-761-1	Sequence 1, Appl
39	25	20.8	556	12	US-10-029-386-12683	Sequence 12683, A
40	25	20.8	582	12	US-10-027-632-88852	Sequence 88852, A
41	25	20.8	582	12	US-10-027-632-88852	Sequence 88852, A
42	25	20.8	582	13	US-10-027-632-88853	Sequence 88853, A
43	25	20.8	582	13	US-10-027-632-88853	Sequence 88853, A
44	25	20.8	2434	14	US-10-198-846-11032	Sequence 11032, A
45	25	20.8	3732	11	US-09-764-891-5825	Sequence 5825, Ap

ALIGNMENTS

RESULT 1

US-09-928-457-73  
; Sequence 73, Application US/09928457  
; Patent No. US2002016403A1  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: DNA, specific proteins and peptides  
; TITLE OF INVENTION: of the Neisseria meningitidis species bacteria, method  
; TITLE OF INVENTION: for obtaining them and their biological application.  
; NUMBER OF SEQUENCES: 99  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30 (OEB)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/928,457  
; FILING DATE: 2001-08-14  
; PRIOR APPLICATION NUMBER: 09/214,759  
; FILING DATE: 199-12-10  
; INFORMATION FOR SEQ ID NO: 73:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 120 base pairs  
; TYPE: nucleotide  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-09-928-457-73

Query Match 100.0%; Score 120; DB 10; Length 120;  
Best Local Similarity 100.0%; Pred. No. 2.9e-36;  
Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CGGTCAGAACAGCGAAGTAAATGAAATCGCTCAGCAGCGAGCTGTGCTCGGACGAAA 60  
DB 1 CGGTCAGAACAGCGAAGTAAATGAAATCGCTCAGCAGCGAGCTGTGCTCGGACGAAA 60

QY 61 ACTCCTTACCGAAGTCTTCTATACCCAGGCTCAATAGCGCTCAAGGAGAGAGCTATCAT 120  
Db 61 ACTCCTTACCGAAGTCTTCTATACCCAGGCTCAATAGCGCTCAAGGAGAGAGCTATCAT 120

## RESULT 2

US-09-928-457-74  
; Sequence 74, Application US/09928457  
; Patent No. US20020164603A1  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: DNA, specific proteins and peptides  
; TITLE OF INVENTION: of the Neisseria meningitidis species bacteria, method  
; TITLE OF INVENTION: for obtaining them and their biological application.  
; NUMBER OF SEQUENCES: 99  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30 (OEB)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/928,457  
; FILING DATE: 2001-08-14  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/214,759  
; FILING DATE: 199-12-10  
; INFORMATION FOR SEQ ID NO: 74:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 120 base pairs  
; TYPE: nucleotide  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-09-928-457-74

Query Match 100.0%; Score 120; DB 10; Length 120;  
Best Local Similarity 100.0%; Pred. No. 2.9e-36;  
Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGTGAGAAACAGGCAAGTATGAAATGCTGAGGACGAGACTGTCTGCGAAGCAAAA 60  
Db 1 CGGTGAGAAACAGGCAAGTATGAAATGCTGAGGACGAGACTGTCTGCGAAGCAAAA 60  
QY 61 ACTCCTTACCGAAGTCTTCTATACCCAGGCTCAATAGCGCTCAAGGAGAGAGCTATCAT 120  
Db 61 ACTCCTTACCGAAGTCTTCTATACCCAGGCTCAATAGCGCTCAAGGAGAGAGCTATCAT 120

## RESULT 3

US-09-938-842A-1646/c  
; Sequence 1646, Application US/09938842A  
; Patent No. US20020160378A1  
; GENERAL INFORMATION:  
; APPLICANT: Harper, Jeff  
; APPLICANT: Kreps, Joel  
; APPLICANT: Wang, Xun  
; APPLICANT: Zhu, Tong  
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING  
; FILE REFERENCE: SCRIPI300-3  
; CURRENT APPLICATION NUMBER: US/09/938,842A  
; CURRENT FILING DATE: 2001-08-24  
; PRIOR APPLICATION NUMBER: US 60/227,866  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: US 60/264,647  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/300,111  
; PRIOR FILING DATE: 2001-06-22  
; NUMBER OF SEQ ID NOS: 5379  
; SEQ ID NO 1646

; LENGTH: 2598  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
US-09-938-842A-1646

Query Match 23.3%; Score 28; DB 10; Length 2598;  
Best Local Similarity 60.5%; Pred. No. 3.2;  
Matches 46; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 10 ACAGGCAAGTAAATGAAATGCTGAGGACGAGACTGTCTGCGAAGCAAAAACCTCTTAC 69  
Db 782 ACTGGAAGAGTACGTAATCTCAGAGGAAGGACCGCTATGAAGGAAAAACCAAGC 723  
QY 70 CGAAGTCTTCTATACC 85  
Db 722 TTACTCTATTAACC 707

## RESULT 4

US-09-923-876-5811/c  
; Sequence 5811, Application US/09923876  
; Patent No. US20020013958A1  
; GENERAL INFORMATION:  
; APPLICANT: Laigudi, Raghunath V.  
; APPLICANT: Kamigaki, Laura Y. (Ito)  
; APPLICANT: Sherman, Bradley K.  
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN SEEDLING  
; FILE REFERENCE: PL-0012-1 CON  
; CURRENT APPLICATION NUMBER: US/09/923,876  
; CURRENT FILING DATE: 2001-08-06  
; PRIOR APPLICATION NUMBER: 09/298,329  
; PRIOR FILING DATE: 1999-04-21  
; PRIOR APPLICATION NUMBER: 60/085,331  
; PRIOR FILING DATE: 1998-05-05  
; NUMBER OF SEQ ID NOS: 6332  
; SOFTWARE: PERL Program  
; SEQ ID NO 5811  
; LENGTH: 261  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No. US20020013958A1 700457772H1  
; NAME/KEY: unsure  
; LOCATION: 17, 70, 114, 131, 186-211, 249  
; OTHER INFORMATION: a, t, c, g, or other  
US-09-923-876-5811

Query Match 23.0%; Score 27.6; DB 9; Length 261;  
Best Local Similarity 70.8%; Pred. No. 2;  
Matches 36; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 29 TGCTGAGGACGAGACTGTCTCGAAGCAAAAACCTCTTACCGAAGCTTC 79  
Db 175 TGCAGCGGCTCGGCTCTGGCCCGCAGCAAAAACCTTGGCGANGCGTTC 125

## RESULT 5

US-10-216-209-1  
; Sequence 1, Application US/10216209  
; Publication No. US20030124634A1  
; GENERAL INFORMATION:  
; APPLICANT: Lam, Joseph S.  
; APPLICANT: Burrows, Lori  
; APPLICANT: Charter, Deborah  
; APPLICANT: De Kievit, Teresa De  
; TITLE OF INVENTION: No. US20030124634A1 Proteins Involved in the Synthesis and Assen  
; TITLE OF INVENTION: of O-Antigen in Pseudomonas Aeruginosa  
; FILE REFERENCE: 6580-167  
; CURRENT APPLICATION NUMBER: US/10/216,209  
; CURRENT FILING DATE: 2002-08-12  
; PRIOR APPLICATION NUMBER: US/09/352,994  
; PRIOR FILING DATE: 2001-05-29

```

; PRIOR APPLICATION NUMBER: US 08/846,762
; PRIOR FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 100
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 24417
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-10-216-209-1

Query Match      22.7%; Score 27.2; DB 14; Length 24417;
Best Local Similarity 61.1%; Pred. No. 14;
Matches 44; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 37 GCACGACTGTCTCGGACGAAACTCTTACCCAACTCTTATACCCAGGCTCAATA 96
Db 14511 GCACGACTGTCTCGGACGAACTTCAAGACGATGTCAGCTTACGGAGGGGTCAA 14570

QY 97 GCGGCTCAAGGA 108
Db 14571 GCGTCTCTCGGA 14582

RESULT 6
US-10-027-632-51299
; Sequence 51299, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 51299
; LENGTH: 525
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-51299

Query Match      22.5%; Score 27; DB 12; Length 525;
Best Local Similarity 53.5%; Pred. No. 4.4;
Matches 54; Conservative 1; Mismatches 46; Indels 0; Gaps 0;

QY 2 GGTCCAGAACGCGCAAGTAAATGAAATGCTTGAGGCACGACTGTGCTGCAACGAAAA 61
Db 39 GGCAGGCTGGAGAACCCACTGACATTCGAAGGCACAGATAATGCTGCTTTGGACAA 98

QY 62 CTCCTTACCGAAGTCTTCTATACCCAGGCTCAATAGCCGCT 102
Db 99 CTGCCAGAGTGTGTCAAACGCGGGGCAAGGGGCTCT 139

RESULT 7
US-10-027-632-51299
; Sequence 51299, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 51299
; LENGTH: 525
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-51299

Query Match      22.5%; Score 27; DB 12; Length 525;
Best Local Similarity 53.5%; Pred. No. 4.4;
Matches 54; Conservative 1; Mismatches 46; Indels 0; Gaps 0;

QY 2 GGTCCAGAACGCGCAAGTAAATGAAATGCTTGAGGCACGACTGTGCTGCAACGAAAA 61
Db 39 GGCAGGCTGGAGAACCCACTGACATTCGAAGGCACAGATAATGCTGCTTTGGACAA 98

QY 62 CTCCTTACCGAAGTCTTCTATACCCAGGCTCAATAGCCGCT 102
Db 99 CTGCCAGAGTGTGTCAAACGCGGGGCAAGGGGCTCT 139

RESULT 8
US-09-960-352-13696
; Sequence 13696, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Mengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 13696
; LENGTH: 394
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (17)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: 58-LIB34-079-Q1-E1-G10
US-09-960-352-13696

Query Match      22.2%; Score 26.6; DB 10; Length 394;
Best Local Similarity 52.2%; Pred. No. 5.6;
Matches 59; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 8 AACACGCGCAAGTAAATGAAATGCTTGAGGCACGACTGTGCTGCAACGAAAACTCCTT 67

```

```
Db 128 AACCGCTCAAGATTGTAATGTTGACTTTGGCAAGAAATGGGATTCATATGAACAAGGATT 187
QY 68 ACCGAAGTCTTCTATACCCAGGCTCAATAGCCGCTCAAGGAGAGAGCTATCAT 120
Db 188 TGGGAATATTTGAACACCAATGATGATGAGGAAGAAATACGTTGGTGTAAACAT 240

RESULT 9
US-10-027-632-54625
; Sequence 54625, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 54625
; LENGTH: 631
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-54625

Query Match 21.8%; Score 26.2; DB 12; Length 631;
Best Local Similarity 56.3%; Pred. No. 9.5;
Matches 49; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 25 AAAATGCTGAGCAGGACTGTGCTGCGAAGAACTCCTTACCGAAGTCTTCTATAC 84
Db 357 AAAATACCTTAGACTGGTAATTGATAAGAACAGAAATTTATTTCTGGAGGCTGGGAAGT 416

QY 85 CCAGGCTCAATAGCCGCTCAAGGAGAG 111
Db 417 CCAAGATCAAAAGCACTGCGAGGTGTG 443

RESULT 10
US-10-027-632-54625
; Sequence 54625, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 54625
; LENGTH: 631
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-54625

Query Match 21.8%; Score 26.2; DB 12; Length 631;
Best Local Similarity 56.3%; Pred. No. 9.5;
Matches 49; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 25 AAAATGCTGAGCAGGACTGTGCTGCGAAGAACTCCTTACCGAAGTCTTCTATAC 84
Db 357 AAAATACCTTAGACTGGTAATTGATAAGAACAGAAATTTATTTCTGGAGGCTGGGAAGT 416

QY 85 CCAGGCTCAATAGCCGCTCAAGGAGAG 111
Db 417 CCAAGATCAAAAGCACTGCGAGGTGTG 443

RESULT 11
US-10-027-632-244583/c
; Sequence 244583, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 244583
; LENGTH: 676
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-244583

Query Match 21.5%; Score 25.8; DB 12; Length 676;
Best Local Similarity 53.5%; Pred. No. 14;
Matches 54; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 2 GGTGAGAAAACAGGCAAGGTAATGAAATAATCCCTGAGGCGAGGACTGTGCTGCGAAGCAAAA 61
Db 258 GGGCAGGCTGGAGGAGCCACTGACATTCGAAGAGGCAGAGATAATGCTGCTTTGGACAA 199

QY 62 CTCCTTACCGAAGTCTTCTATACCCAGGCTCAATAGCCGCT 102
Db 198 CTGCCAGAGTGTGTGACACCAGCGGGCCAAAGGGGCTCT 158

RESULT 12
US-10-027-632-244584/c
```

```
; Sequence 244584, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 244584
; LENGTH: 676
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-244584

Query Match 21.5%; Score 25.8; DB 12; Length 676;
Best Local Similarity 53.5%; Pred. No. 14;
Matches 54; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 2 GGTGAGAAACAGGCAAGTAAATGAAATGCTGAGGACGAGCTGTGTCGGAACGAAAA 61
Db 258 GGCAGGCTGGAGGAGCCACTGACATTGCAAGAGGCACAGATAATGCTGCTTTGGACAA 199

QY 62 CTCCTTACCGAAGTCTTCTATACCCAGGCTCAATAGCCGCT 102
Db 198 CTGCCAGAGTGTGCAACAGCCGGGGCCCAAGGGGCTCT 158

RESULT 14
US-10-027-632-244584/c
; Sequence 244584, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 244584
; LENGTH: 676
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-244584

Query Match 21.5%; Score 25.8; DB 13; Length 676;
Best Local Similarity 53.5%; Pred. No. 14;
Matches 54; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 2 GGTGAGAAACAGGCAAGTAAATGAAATGCTGAGGACGAGCTGTGTCGGAACGAAAA 61
Db 258 GGCAGGCTGGAGGAGCCACTGACATTGCAAGAGGCACAGATAATGCTGCTTTGGACAA 199

QY 62 CTCCTTACCGAAGTCTTCTATACCCAGGCTCAATAGCCGCT 102
Db 198 CTGCCAGAGTGTGCAACAGCCGGGGCCCAAGGGGCTCT 158

RESULT 15
US-10-027-632-8784
; Sequence 8784, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
```

```
/ PRIOR FILING DATE: 2000-07-12
/ PRIOR APPLICATION NUMBER: US 60/198,676
/ PRIOR FILING DATE: 2000-04-20
/ PRIOR APPLICATION NUMBER: US 60/193,483
/ PRIOR FILING DATE: 2000-03-29
/ PRIOR APPLICATION NUMBER: US 60/185,218
/ PRIOR FILING DATE: 2000-02-24
/ PRIOR APPLICATION NUMBER: US 60/167,363
/ PRIOR FILING DATE: 1999-11-23
/ PRIOR APPLICATION NUMBER: US 60/156,358
/ PRIOR FILING DATE: 1999-09-28
/ PRIOR APPLICATION NUMBER: US 60/146,002
/ PRIOR FILING DATE: 1999-08-09
/ NUMBER OF SEQ ID NOS: 325720
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 8784
/ LENGTH: 876
/ TYPE: DNA
/ ORGANISM: Human
US-10-027-632-8784

Query Match      21.3%; Score 25.6; DB 12; Length 876;
Best Local Similarity 57.5%; Pred. No. 18;
Matches 46; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY      8 AAACAGGCAAGGTATGAAATCCCTGAGCGGACTGTGCTGCGAAGCAAACTCGTT 67
Db      446 AAACCTTGCAAGACAGACAGAAAGGGCTTACGTATGGCAATATTAGATAAAACTATTT 505

QY      68 ACCGAAGTCTTCTATACCCA 87
Db      506 TCTAGGATAATATACCCCTA 525
```

Search completed: November 15, 2003, 08:31:54  
Job time : 74.0536 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 14, 2003, 23:56:24 ; Search time 1190.78 Seconds  
(without alignments)  
9241.609 Million cell updates/sec

Title: US-09-928-457-77  
Perfect score: 269  
Sequence: 1 CGAGGATGAAATCGTTATT.....GCAAGGGTATTATCAACCG 269

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues  
Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.ov.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.hum.\*
- 18: em.in.\*
- 19: em.mu.\*
- 20: em.om.\*
- 21: em.or.\*
- 22: em.ov.\*
- 23: em.pat.\*
- 24: em.ph.\*
- 25: em.pl.\*
- 26: em.ro.\*
- 27: em.sts.\*
- 28: em.un.\*
- 29: em.vi.\*
- 30: em.htg.hum.\*
- 31: em.htg.inv.\*
- 32: em.htg.other.\*
- 33: em.htg.mus.\*
- 34: em.htg.pln.\*
- 35: em.htg.rtd.\*
- 36: em.htg.nam.\*
- 37: em.htg.vrt.\*
- 38: em.sy.\*
- 39: em.htgo.hum.\*
- 40: em.htgo.mus.\*
- 41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	269	100.0	269	1	AF169451	AF169451 Neisseria
2	269	100.0	269	6	A68906	A68906 Sequence 77
3	269	100.0	269	6	BD063006	BD063006 DNA and s
4	267.4	99.4	349061	1	NMA222491	AL162753 Neisseria
5	264.2	98.2	858	6	AX024058	AX024058 Sequence
6	246.6	91.7	858	6	AX024111	AX024111 Sequence
C 7	121.8	45.3	270	6	A68917	A68917 Sequence 88
C 8	121.8	45.3	270	6	BD063017	BD063017 DNA and s
C 9	121.8	45.3	271	1	AF169464	AF169464 Neisseria
C 10	110.8	41.2	270	1	AF169428	AF169428 Neisseria
11	67.8	25.2	4884	1	HGDIIM	X55141 H.giganteus
12	42.6	15.8	164490	2	AC079077	AC079077 Homo sapi
13	42.2	15.7	63215	2	AC061968	AC061968 Homo sapi
14	41	15.2	185539	2	AC018349	AC018349 Homo sapi
15	39.4	14.6	125746	9	AC010279	AC010279 Homo sapi
16	39.4	14.6	142715	9	AC008387	AC008387 Homo sapi
C 17	38.8	14.4	254912	2	AC123482	AC123482 Rattus no
C 18	38.8	14.4	333352	2	AC113855	AC113855 Rattus no
19	38.8	14.4	333352	2	AC125577	AC125577 Rattus no
C 20	38.2	14.2	123098	9	AC008799	AC008799 Homo sapi
21	38.2	14.2	151031	9	AC008923	AC008923 Homo sapi
C 22	38	14.1	187166	2	AC138297	AC138297 Mus muscu
23	37.6	14.0	110000	2	AC117022_1	Continuation (2 of
24	37.6	14.0	121922	2	AP000576	AP000576 Homo sapi
25	37.6	14.0	161582	9	AP001636	AP001636 Homo sapi
26	37.6	14.0	161586	9	AC090309	AC090309 Homo sapi
27	37.6	14.0	169670	9	AC023908	AC023908 Homo sapi
C 28	37.6	14.0	239844	2	AC134274	AC134274 Rattus no
C 29	37.4	13.9	144201	10	AC087218	AC087218 Rattus no
C 30	37.4	13.9	227796	2	AC131418	AC131418 Rattus no
C 31	37.4	13.9	248922	2	AC129700	AC129700 Rattus no
32	37.2	13.8	44298	3	U58751	U58751 Caenorhabdi
C 33	37.2	13.8	143092	3	CEV17G7B	AL023828 Caenorhab
34	37	13.8	192969	10	AC122373	AC122373 Mus muscu
35	36.8	13.7	2578	3	AY051762	AY051762 Drosophil
36	36.8	13.7	3170	3	AB043434	AB043434 Drosophil
C 37	36.8	13.7	20175	2	AC019990	AC019990 Drosophil
C 38	36.8	13.7	55307	2	AC014684	AC014684 Drosophil
C 39	36.8	13.7	150391	2	AC011764	AC011764 Drosophil
C 40	36.8	13.7	162345	3	AC006415	AC006415 Drosophil
C 41	36.8	13.7	170498	3	AC008094	AC008094 Drosophil
C 42	36.8	13.7	186695	3	AC013351	AC013351 Drosophil
C 43	36.8	13.7	244757	3	AE003678	AE003678 Drosophil
C 44	36.8	13.7	298116	3	AE003782	AE003782 Drosophil
45	36.6	13.6	179607	2	AC102778	AC102778 Mus muscu

ALIGNMENTS

RESULT 1  
AF169451  
LOCUS AF169451  
DEFINITION Neisseria meningitidis strain Z2491 clone Cm040 unknown sequence.  
ACCESSION AF169451  
VERSION AF169451.1  
KEYWORDS GI:9754659  
SOURCE Neisseria meningitidis  
ORGANISM Neisseria meningitidis  
Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
Neisseriaceae; Neisseria.  
1 (bases 1 to 269)  
Perrin A., Nassif, X. and Tinsley, C.R.  
Identification of regions of the chromosome of Neisseria meningitidis and Neisseria gonorrhoeae which are specific to

```

pathogenic Neisseriae
Unpublished
REFERENCE 2 (bases 1 to 269)
AUTHORS Perrin A., Nassif, X. and Tinsley, C.R.
TITLE Direct Submission
JOURNAL Submitted (16-JUL-1999) Necker Medicine Faculty, INSERM U411, 156
rue de Vaugirard, Paris 75015, France
FEATURES
    source
        Location/Qualifiers
            1..269
                /organism="Neisseria meningitidis"
                /mol_type="genomic DNA"
                /strain="Z2491"
                /db_xref="taxon:487"
                /clone="Cm040"
BASE COUNT 95 a 41 c 65 g 68 t
ORIGIN
    Query Match 100.0%; Score 269; DB 1; Length 269;
    Best Local Similarity 100.0%; Pred. No. 9e-60;
    Matches 269; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGAGCATAAATCGTTATTAAAGATAATGGTATAGGAACGAGCTTCGATGAATCAATG 60
Db 1 CGGAGCATAAATCGTTATTAAAGATAATGGTATAGGAACGAGCTTCGATGAATCAATG 60
QY 61 ATTTTATTGAGATCGGTTCGGAACAGAGGGAAGGAAACAAAGCTCCCGTCGGAA 120
Db 61 ATTTTATTGAGATCGGTTCGGAACAGAGGGAAGGAAACAAAGCTCCCGTCGGAA 120
QY 121 GAATTCACCGGTAAAGGCGCTTGGTAAATTTGGCATTATTCGGGCTTGGCAACAAA 180
Db 121 GAATTCACCGGTAAAGGCGCTTGGTAAATTTGGCATTATTCGGGCTTGGCAACAAA 180
QY 181 TTGAAATTTCTACTATCCAGGGAACGAAAGGTTACTTTTACTTTTGGATTATGACAGA 240
Db 181 TTGAAATTTCTACTATCCAGGGAACGAAAGGTTACTTTTACTTTTGGATTATGACAGA 240
QY 241 TTCGAAGAAGCAAGGTTATTATCAACCG 269
Db 241 TTCGAAGAAGCAAGGTTATTATCAACCG 269

RESULT 3
BD063006 269 bp DNA linear PAT 27-AUG-2002
LOCUS DNA and specific proteins or peptides of the Neisseria meningitidis
DEFINITION species bacteria, method for obtaining them and their biological
ACCESSION BD063006
VERSION BD063006.1 GI:22608609
KEYWORDS JP 2001504684-A/68.
SOURCE unidentified
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 269)
AUTHORS Nassif, X., Tinsley, C., Achtman, M., Ruelle, J., Vinals, C. and
Marker, P.
TITLE DNA and specific proteins or peptides of the Neisseria meningitidis
JOURNAL species bacteria, method for obtaining them and their biological
applications
COMMENT PN JP 2001504684-A/68
PD 10-APR-2001
PF 11-JUL-1997 JP 1998050585
PR 12-JUL-1996 FR 96/08768
PI XAVIER NASSIF, COLIN TINSLEY, MARK ACHTMAN, JEAN LOUIS RUELLE, PI
CARLA VINALS,
PI PETRA MERKER
PC Cl2N15/31, C07K14/22, C07K16/12, A61K39/095, C12Q1/68, G01N33/53 CC
Strandedness: Single;
CC Topology: Linear;
FH Key Location/Qualifiers.
FEATURES
    source
        1..269
            /organism="unidentified"
            /mol_type="genomic DNA"
            /db_xref="taxon:32644"
BASE COUNT 95 a 41 c 65 g 68 t
ORIGIN
    Query Match 100.0%; Score 269; DB 6; Length 269;
    Best Local Similarity 100.0%; Pred. No. 9e-60;
    Matches 269; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGAGCATAAATCGTTATTAAAGATAATGGTATAGGAACGAGCTTCGATGAATCAATG 60
Db 1 CGGAGCATAAATCGTTATTAAAGATAATGGTATAGGAACGAGCTTCGATGAATCAATG 60
QY 61 ATTTTATTGAGATCGGTTCGGAACAGAGGGAAGGAAACAAAGCTCCCGTCGGAA 120
Db 61 ATTTTATTGAGATCGGTTCGGAACAGAGGGAAGGAAACAAAGCTCCCGTCGGAA 120
QY 121 GAATTCACCGGTAAAGGCGCTTGGTAAATTTGGCATTATTCGGGCTTGGCAACAAA 180
Db 121 GAATTCACCGGTAAAGGCGCTTGGTAAATTTGGCATTATTCGGGCTTGGCAACAAA 180
QY 181 TTGAAATTTCTACTATCCAGGGAACGAAAGGTTACTTTTACTTTTGGATTATGACAGA 240
Db 181 TTGAAATTTCTACTATCCAGGGAACGAAAGGTTACTTTTACTTTTGGATTATGACAGA 240
QY 241 TTCGAAGAAGCAAGGTTATTATCAACCG 269
Db 241 TTCGAAGAAGCAAGGTTATTATCAACCG 269

RESULT 2
A68906 269 bp DNA linear PAT 06-MAY-1999
LOCUS Sequence 77 from Patent WO9802547.
DEFINITION
ACCESSION A68906
VERSION A68906.1 GI:4759825
SOURCE unidentified
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 269)
AUTHORS Nassif, X., Tinsley, C., Achtman, M., Ruelle, J., Vinals, C. and
Marker, P.
TITLE DNA AND SPECIFIC PROTEINS OR PEPTIDES OF THE NEISSERIA MENINGITIDIS
JOURNAL SPECIES BACTERIA, METHOD FOR OBTAINING THEM AND THEIR BIOLOGICAL
APPLICATIONS
COMMENT Patent: WO 9802547-A 77 22-JAN-1998;
INST NAT SANTE RECH MED (FR)
Other publication FR 2751000 19980116.
FEATURES
    source
        Location/Qualifiers
            1..269
                /organism="unidentified"
                /mol_type="genomic DNA"
                /db_xref="taxon:32644"
BASE COUNT 95 a 41 c 65 g 68 t
ORIGIN
    Query Match 100.0%; Score 269; DB 6; Length 269;
    Best Local Similarity 100.0%; Pred. No. 9e-60;
    Matches 269; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 CGGAGCATAAATCGTTATTAAAGATAATGGTATAGGAACGAGCTTCGATGAATCAATG 60
Db 1 CGGAGCATAAATCGTTATTAAAGATAATGGTATAGGAACGAGCTTCGATGAATCAATG 60
QY 61 ATTTTATTGAGATCGGTTCGGAACAGAGGGAAGGAAACAAAGCTCCCGTCGGAA 120
Db 61 ATTTTATTGAGATCGGTTCGGAACAGAGGGAAGGAAACAAAGCTCCCGTCGGAA 120
QY 121 GAATTCACCGGTAAAGGCGCTTGGTAAATTTGGCATTATTCGGGCTTGGCAACAAA 180
Db 121 GAATTCACCGGTAAAGGCGCTTGGTAAATTTGGCATTATTCGGGCTTGGCAACAAA 180
QY 181 TTGAAATTTCTACTATCCAGGGAACGAAAGGTTACTTTTACTTTTGGATTATGACAGA 240
Db 181 TTGAAATTTCTACTATCCAGGGAACGAAAGGTTACTTTTACTTTTGGATTATGACAGA 240
QY 241 TTCGAAGAAGCAAGGTTATTATCAACCG 269
Db 241 TTCGAAGAAGCAAGGTTATTATCAACCG 269

RESULT 3
BD063006 269 bp DNA linear PAT 27-AUG-2002
LOCUS DNA and specific proteins or peptides of the Neisseria meningitidis
DEFINITION species bacteria, method for obtaining them and their biological
ACCESSION BD063006
VERSION BD063006.1 GI:22608609
KEYWORDS JP 2001504684-A/68.
SOURCE unidentified
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 269)
AUTHORS Nassif, X., Tinsley, C., Achtman, M., Ruelle, J., Vinals, C. and
Marker, P.
TITLE DNA and specific proteins or peptides of the Neisseria meningitidis
JOURNAL species bacteria, method for obtaining them and their biological
applications
COMMENT PN JP 2001504684-A/68
PD 10-APR-2001
PF 11-JUL-1997 JP 1998050585
PR 12-JUL-1996 FR 96/08768
PI XAVIER NASSIF, COLIN TINSLEY, MARK ACHTMAN, JEAN LOUIS RUELLE, PI
CARLA VINALS,
PI PETRA MERKER
PC Cl2N15/31, C07K14/22, C07K16/12, A61K39/095, C12Q1/68, G01N33/53 CC
Strandedness: Single;
CC Topology: Linear;
FH Key Location/Qualifiers.
FEATURES
    source
        1..269
            /organism="unidentified"
            /mol_type="genomic DNA"
            /db_xref="taxon:32644"
BASE COUNT 95 a 41 c 65 g 68 t
ORIGIN
    Query Match 100.0%; Score 269; DB 6; Length 269;
    Best Local Similarity 100.0%; Pred. No. 9e-60;
    Matches 269; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGAGCATAAATCGTTATTAAAGATAATGGTATAGGAACGAGCTTCGATGAATCAATG 60
Db 1 CGGAGCATAAATCGTTATTAAAGATAATGGTATAGGAACGAGCTTCGATGAATCAATG 60
QY 61 ATTTTATTGAGATCGGTTCGGAACAGAGGGAAGGAAACAAAGCTCCCGTCGGAA 120
Db 61 ATTTTATTGAGATCGGTTCGGAACAGAGGGAAGGAAACAAAGCTCCCGTCGGAA 120
QY 121 GAATTCACCGGTAAAGGCGCTTGGTAAATTTGGCATTATTCGGGCTTGGCAACAAA 180
Db 121 GAATTCACCGGTAAAGGCGCTTGGTAAATTTGGCATTATTCGGGCTTGGCAACAAA 180
QY 181 TTGAAATTTCTACTATCCAGGGAACGAAAGGTTACTTTTACTTTTGGATTATGACAGA 240
Db 181 TTGAAATTTCTACTATCCAGGGAACGAAAGGTTACTTTTACTTTTGGATTATGACAGA 240
QY 241 TTCGAAGAAGCAAGGTTATTATCAACCG 269
Db 241 TTCGAAGAAGCAAGGTTATTATCAACCG 269

RESULT 2
A68906 269 bp DNA linear PAT 06-MAY-1999
LOCUS Sequence 77 from Patent WO9802547.
DEFINITION
ACCESSION A68906
VERSION A68906.1 GI:4759825
SOURCE unidentified
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 269)
AUTHORS Nassif, X., Tinsley, C., Achtman, M., Ruelle, J., Vinals, C. and
Marker, P.
TITLE DNA AND SPECIFIC PROTEINS OR PEPTIDES OF THE NEISSERIA MENINGITIDIS
JOURNAL SPECIES BACTERIA, METHOD FOR OBTAINING THEM AND THEIR BIOLOGICAL
APPLICATIONS
COMMENT Patent: WO 9802547-A 77 22-JAN-1998;
INST NAT SANTE RECH MED (FR)
Other publication FR 2751000 19980116.
FEATURES
    source
        Location/Qualifiers
            1..269
                /organism="unidentified"
                /mol_type="genomic DNA"
                /db_xref="taxon:32644"
BASE COUNT 95 a 41 c 65 g 68 t
ORIGIN
    Query Match 100.0%; Score 269; DB 6; Length 269;
    Best Local Similarity 100.0%; Pred. No. 9e-60;
    Matches 269; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```



```

QY 121 GAATTCACCGGTAAAAAGCGCTTGGTAAATGGCATTATTCGGCTTGGCAACAAA 180
Db 121 GAATTCACCGGTAAAAAGCGCTTGGTAAATGGCATTATTCGGCTTGGCAACAAA 180

QY 181 TTGAATTTCTACTATCCAGGGAACGAAAGGGTTACTTTTACTTTGGATTATGCAGAGA 240
Db 181 TTGAATTTCTACTATCCAGGGAACGAAAGGGTTACTTTTACTTTGGATTATGCAGAGA 240

QY 241 TTCGAGAACCAAGGTTATTATCAACCG 269
Db 241 TTCGAGAACCAAGGTTATTATCAACCG 269

RESULT 4
NMA22491
LOCUS NMA22491 349061 bp DNA linear BCT 02-SEP-2002
DEFINITION Neisseria meningitidis serogroup A strain 22491 complete genome;
segment 2/7.
ACCESSION AL162753 AL157959
VERSION AL162753.2 GI:7379120
KEYWORDS
SOURCE
ORGANISM Neisseria meningitidis Z2491
Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
Neisseriaceae; Neisseria.
REFERENCE
1 (bases 1 to 349061)
Parkhill, J., Achtman, M., James, K.D., Bentley, S.D., Churcher, C.,
Klee, S.R., Morelli, G., Basham, D., Brown, D., Chillingworth, T.,
Davies, R.M., Davis, P., Devlin, K., Felwell, T., Hamlin, N.,
Kholodny, S., Jagers, K., Leather, S., Mungall, K.,
Quail, M.A., Rajandream, M.A., Rutherford, K.M., Simmonds, M.,
Skellton, J., Whitehead, S., Spratt, B.G. and Barrell, B.G.
Complete DNA sequence of a serogroup A strain of Neisseria
meningitidis Z2491
NATURE 404 (6777), 502-506 (2000)
JOURNAL
MEDLINE 2022556
PUBMED 10761919
REFERENCE
2 (bases 1 to 349061)
Parkhill, J.
DIRECT SUBMISSION
Submitted (30-MAR-2000) Submitted on behalf of the Neisseria
sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk
NOTES
Details of N. meningitidis sequencing at the Sanger Centre are
available on the World Wide Web.
(URL, http://www.sanger.ac.uk/projects/N\_meningitidis/).
FEATURES
Source
1..349061
/organism="Neisseria meningitidis Z2491"
/mol_type="genomic DNA"
/strain="Z2491"
/db_xref="taxon:122587"
/note="serogroup: A"
complement(24..206)
/note="ATR repeat; hmms hit to HMM ATR (1 - 183), score:
310.39"
/label="ATR"
RBS 209..212
Gene 220..681
CDS 220..681
/gene="NMA0368"
/note="NMA0368"
/note="NMA0368", probable integral membrane protein, len:
153 aa; contains four probable transmembrane domains"
/codon_start=1
/transl_table=11
/product="putative integral membrane protein"
/protein_id="CAB83669.1"
/db_xref="GI:7379121"
/translation="MQEQNRKSSFPFIVMLLVSVLWIASLNVAFYLGHSMEGLTV

```

```

LILGSIFASLDIRYCAVYVNWLAALVLLALRKKVVPVHAAFWGLALVAFSVKAVVY
DEAGNTSDIVRYGAGFVLYAAFAVASIGTFAGNKKERKAASAADGTKNDY"
671..1492
/gene="hemK"
671..1492
/gene="hemK"
/note="NMA0369, hemK, HemK protein, len: 273 aa; similar
to e.g. HEMK_ECOLI P37186 HEMK protein (277 aa), fasta
scores; E(): 0.423% identity in 279 aa overlap. Contains
PS00092 N-6 Adenine-specific DNA methylases signature"
/codon_start=1
/transl_table=11
/product="HemK protein"
/protein_id="CAB83670.1"
/db_xref="GI:7379122"
/translation="MTDFKMLGSKLPKNEARMLLQVSEYRVTQLLRGGEEMPDEV
RQRADLAQRRLNGEVPAYILGAREFYGRFTVNPFSVLIRPEPTEHLVEAVLAKLPEN
GRVNDLTGSGAVATVVALRPPDAFVRASDISPEALTARKNAADLGARVEFAYGSWF
DTMPSGKMDLIIVSNPPYIENGKHLISQDGLRPEPOLATDFDSCGLSCRTLAQGAP
DRLAEGGFLLEHGFDDQAAVRAVGLAENGSGSVETLPDLAGLDRTVLGKYNKHUK"
831..840
/gene="hemK"
/note="Core DNA uptake sequence: gcgctgtgaa"
/label=DUS
1162..1171
/gene="hemK"
/note="Core DNA uptake sequence: gcgctgtgaa"
/label=DUS
1187..1207
/gene="hemK"
/note="PS00092 N-6 Adenine-specific DNA methylases
signature"
complement(1279..1288)
/note="Core DNA uptake sequence: gcgctgtgaa"
/label=DUS
1578..1568
/gene="NMA0370"
1578..2966
/gene="NMA0370"
/note="NMA0370"
/note="NMA0370, probable integral membrane protein, len:
462 aa; similar to hypothetical proteins e.g. Y325_HABIN
P44640 hypothetical protein H10325 (450 aa), fasta scores;
E(): 0.498% identity in 464 aa overlap"
/codon_start=1
/transl_table=11
/product="putative integral membrane protein"
/protein_id="CAB83671.1"
/db_xref="GI:7379123"
/translation="MNAVVAIVMLVLSLRVHVLSLTIGAFVGAVAGMPLQNTA
DAAGQVSQAGIIPVKNKLGSGAKIALSYAMLGAPAMAITHSGLPOOLAGAVRKLNR
GMPDSVRSCEGVKVLKLLSILVGMMSQNIPIHIAFIPNIVPPELLVFNRLKIDR
RLIACVITFLVTTVTFPLPYGGAIFLINEILLGNHSAAPOLDVKNVNNVAAWAIPL
GMIALLLAFHYRKPRLYQSNNDAGNDAANRPQPSAYRSLAAVAIVCFAIQL
MYEDSLVGLMGFVFMGLGVNRDKANDVFGEIKMMANVGFIIMAAQGFVAVVNA
TGHITQIVESSMAIFGNSKGMALMLVGLLVMTGIGSSFSFTLPIAAYVPLCVGL
GFSPLATVAIVGTAGALGDAGSPASDSTLGTPTWGLNADGQHDHRSVITFIHYNIP
LLIAGMTAAWVL"
2967..3191
/gene="slyX"
2967..3191
/gene="slyX"
/note="NMA0371, slyX, slyX protein homolog, len: 74 aa;
similar to slyX_HABIN P44759 slyX protein homolog (73 aa),
fasta scores; E(): 0.0017, 33.8% identity in 74 aa
overlap, and slyX_ECOLI P30857 slyX protein (72 aa), fasta
scores; E(): 0.24, 32.4% identity in 68 aa overlap"
/codon_start=1
/transl_table=11
/product="slyX protein homolog"
/protein_id="CAB83672.1"

```







Herpetosiphon giganteus strain Hpa2  
Gene 106 (1), 87-92 (1991)  
MEDLINE 92039068  
PUBMED 1937045  
COMMENT See also X55137-X55143  
FEATURES Location/Qualifiers  
Source 1..4884  
/organism="Herpetosiphon aurantiacus"  
/mol\_type="genomic DNA"  
/strain="Hpa2"  
/db\_xref="taxon:65"  
884..1300  
/gene="orf 15"  
884..1300  
/gene="orf 15"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="CAA38940.1"  
/db\_xref="GI:48772"  
/db\_xref="SWISS-PROT:P25279"  
/translation="MWTAGYLTREAIGLCRNDVAHEAGTNNVQIMRIEKGIDTR  
GSLLSVRAVFNFAEHIQFLMLVATEEDGNLAIWINRETTGAIDEFIADVKD  
NKVEALKLIQLEALDPSLLGLVQSLDRNR"  
1749..2813  
/gene="hgidiim"  
1749..2813  
/gene="hgidiim"  
/codon\_start=1  
/transl\_table=1  
/product="methyltransferase"  
/protein\_id="CAA38941.1"  
/db\_xref="GI:48773"  
/db\_xref="SWISS-PROT:P25265"  
/translation="MVGAVIDLFCGVGLTHGILLGSGVLACIDNDPCKYAYEONN  
RTPFKSIEVDGRELNALPNKHLVGCAPQDPQYTKRSRTGKTKWLLTEFS  
RLRIEPEPDIEMNPEVFNTRGEVFNFIQSLQGLYHVSHVPCDVPICPOOR  
DLVLFAAKGVGKIKIPPTPENRYTRDVIGSLATNYSGGHWEGDSWHAASRLIED  
NLRIIOHSPGVGTWAPSELIAECHKESGESYGVGRMWDKVAPIITQCNGYG  
NGRIHGEODRA.SLREALLQTFPSYQFAPGQLKFTVSRQIGNVPAVALGRVIA  
KSTKRLGLEHQRVRILL"  
2782..4617  
/gene="orf 68"  
2782..4617  
/gene="orf 68"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="CAA38942.1"  
/db\_xref="GI:48774"  
/db\_xref="SWISS-PROT:P25280"  
/translation="MDSGYEISLNTLNHLSEGLYSNLFYKLKLVANAMDATE  
VSNIRQDQIVIQDNGISIEDANTKFLRIGHQKREDSANTISGRHYMGRKIGILA  
IFGIANIAEYVCKDGVPHGFTIHKGDIERGISSDVTLYRPSVPCEALSIESTKII  
LREIKSSIGNAEKLTLDLARRFTIINNNSFVLIINTPINDNRDYLKQVFLWYL  
GSESSKXANFTFKLKSFEITNLGDGSGMTVKGWIGTVRPSDIPNQPYHYFAHG  
KMLQEDILDI DAGRYOTIIGIEADPFMDSDDEDDIITSNEKRIKOTDPYLLKLE  
YKIDIMRVIASTWTKLREYPSNPKKEVNDSSLSKDSNSEKENTASSDSSTATE  
NASSDSSTATENSDGEVDNSFFDDDIPEPSPKQEIITAFREMKLVKQSN  
IPNQKNIILYDIQAAVAYKGTSKACIVMLGAILEGVMGICITORTDVLVLIJALPO  
VPKPLSDLPKPKPADRYVLAQYITFTSFPODCKEILBLCVGTNKLGVLDILOVRN  
SIHPSVLKDMQLAFNFHQSAVGVIKHLINLVILNWPSP"  
BASE COUNT 1500 a 882 c 1055 g 1447 t  
ORIGIN

Query Match 25.2%; Score 67.8; DB 1; Length 4884;  
Best Local Similarity 53.6%; Pred. No. 2.4e-07;  
Matches 141; Conservative 0; Mismatches 122; Indels 0; Gaps 0;

QY 7 ATAAATCGTTTAAAGATATGTTATAGGACGAGCTTCGATGAATCAATGATTTT 66  
|||||  
Db 2936 ATCAATAGTCATCCAGGATAATGTTATAGGATGTCAATTGAGGATGCTAATAAT 2995  
|||||

QY 67 ATTGAGATCGTCCGACAGAGGAGAGAAACAAACCTCCCGCGCGAAGATTC 126  
|||||

2996 TCTTACGAATTGGCATCAAAAGAGAGAGATTACGTAACACTATAGTGGACGCATG 3055  
QY 127 CAACGGTAAAAAGCCTTGGTAATTCGCATTTATTCGGGTTGGCAACAAATGAAA 186  
|||||  
Db 3056 TTATGGCAGGAAAGGATCGGTATCTACTGCAATATTTGGTATAGCAATATTGCTGAG 3115  
|||||  
QY 187 TTCTACTATCCAGGAAACGAAAGGTTACTTTTACTTTGGATTATGCAGAGATTCGAA 246  
|||||  
Db 3116 TTTATCTTGTGAAGGATGAGTACCCCATGGATTTATAATTATAGGCGGATATAGAAA 3175  
|||||  
QY 247 GAAGCAAGGTTATTTATCAACCG 269  
|||||  
Db 3176 GAGGATAAGTAGCGATGTACG 3198  
|||||

RESULT 12  
AC079077  
LOCUS 164490 bp DNA linear HTG 24-AUG-2002  
DEFINITION Homo sapiens chromosome 5 clone RP11-424A10 map 5, WORKING DRAFT  
SEQUENCE, 42 unordered pieces.  
AC079077  
AC079077.2 GI:11178110  
HTG; HTGS PHASE1; HTGS\_DRAFT.  
VERSION  
KEYWORDS Homo sapiens (human)  
SOURCE  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 164490)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
Anderson,S., Barria,N., Bastien,V., Beda,F., Boguslavsky,L.,  
Bouckhalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,  
Chepel,Y., Colangelo,M., Collins,S., Collinmore,A., Cooke,P.,  
Dearlano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,  
Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M.,  
Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L.,  
Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., LaRocque,K.,  
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G.,  
Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McKernan,K.,  
McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,  
Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T.,  
O'Donnell,P., Pisanic,C., Pollara,V., Raymond,C., Rieback,M., Riley,R.,  
Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P.,  
Sougnuez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,  
Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,  
Zimmer,A. and Zody,M.  
Direct Submission  
Submitted (17-AUG-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 164490)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
Anderson,S., Barria,N., Bastien,V., Beda,F., Boguslavsky,L.,  
Bouckhalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,  
Chepel,Y., Colangelo,M., Collins,S., Collinmore,A., Cooke,P.,  
Dearlano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,  
Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M.,  
Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L.,  
Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., LaRocque,K.,  
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G.,  
Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McKernan,K.,  
McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,  
Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T.,  
O'Donnell,P., Pisanic,C., Pollara,V., Raymond,C., Rieback,M., Riley,R.,  
Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P.,  
Sougnuez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,

Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A.,  
 Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J.,  
 Zimmer, A., and Zody, M.  
 Direct Submission  
 Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Nov 15, 2000 this sequence version replaced gi:9838006.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR  
 Web site: <http://www-seq.wi.mit.edu>  
 Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
 Project Information  
 Center project name: L10780  
 Center clone name: 424\_A\_10

Summary Statistics  
 Sequencing vector: Plasmid; n/a; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 140058 bases at least Q40  
 Consensus quality: 152620 bases at least Q30  
 Consensus quality: 157604 bases at least Q20  
 Insert size: 183000; agarose-fp  
 Insert size: 160390; sum-of-contigs.  
 Quality coverage: 2.7 in Q20 bases; agarose-fp  
 Quality coverage: 3.1 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently  
 consists of 42 contigs. The true order of the pieces  
 is not known and their order in this sequence record is  
 arbitrary. Gaps between the contigs are represented as  
 runs of N, but the exact sizes of the gaps are unknown.  
 This record will be updated with the finished sequence  
 as soon as it is available and the accession number will  
 be preserved.

1 379: contig of 379 bp in length  
 380 479: gap of 100 bp  
 480 15053: contig of 14574 bp in length  
 15054 15153: gap of 100 bp  
 15154 15720: contig of 567 bp in length  
 15721 15820: gap of 100 bp  
 15821 17135: contig of 1315 bp in length  
 17136 17235: gap of 100 bp  
 17236 18343: contig of 1108 bp in length  
 18344 18443: gap of 100 bp  
 18444 19570: contig of 1127 bp in length  
 19571 19670: gap of 100 bp  
 19671 21870: contig of 2200 bp in length  
 21871 21970: gap of 100 bp  
 21971 23558: contig of 1588 bp in length  
 23559 23658: gap of 100 bp  
 23659 24734: contig of 1076 bp in length  
 24735 24834: gap of 100 bp  
 24835 26227: contig of 1393 bp in length  
 26228 26327: gap of 100 bp  
 26328 28306: contig of 1979 bp in length  
 28307 28405: gap of 100 bp  
 28407 29823: contig of 1423 bp in length  
 29830 29923: gap of 100 bp  
 29930 31365: contig of 1436 bp in length  
 31366 31465: gap of 100 bp  
 31466 33041: contig of 1576 bp in length  
 33042 33141: gap of 100 bp  
 33142 35033: contig of 1898 bp in length  
 35034 35133: gap of 100 bp  
 35140 37112: contig of 1973 bp in length  
 37113 37213: gap of 100 bp  
 37213 38733: contig of 1521 bp in length  
 38734 38833: gap of 100 bp  
 38834 41123: contig of 2290 bp in length

41124 41223: gap of 100 bp  
 41224 43464: contig of 2241 bp in length  
 43465 43564: gap of 100 bp  
 43565 46245: contig of 2681 bp in length  
 46246 46345: gap of 100 bp  
 46346 49174: contig of 2829 bp in length  
 49175 49274: gap of 100 bp  
 49275 51190: contig of 1916 bp in length  
 51191 52290: gap of 100 bp  
 52291 53683: contig of 2403 bp in length  
 53684 53783: gap of 100 bp  
 53784 57615: contig of 3822 bp in length  
 57616 57715: gap of 100 bp  
 57716 60918: contig of 3203 bp in length  
 60919 61019: gap of 100 bp  
 61020 63664: contig of 2645 bp in length  
 63665 63763: gap of 100 bp  
 63764 67175: contig of 3412 bp in length  
 67176 67275: gap of 100 bp  
 67276 70782: contig of 3507 bp in length  
 70783 75448: contig of 4565 bp in length  
 75449 75548: gap of 100 bp  
 75549 79599: contig of 4051 bp in length  
 79600 79699: gap of 100 bp  
 79700 84596: contig of 4897 bp in length  
 84597 84696: gap of 100 bp  
 84697 89528: contig of 4832 bp in length  
 89529 89628: gap of 100 bp  
 89629 94192: contig of 4564 bp in length  
 94193 94292: gap of 100 bp  
 94293 99758: contig of 5466 bp in length  
 99759 105474: contig of 5616 bp in length  
 105475 105574: gap of 100 bp  
 105575 111265: contig of 5591 bp in length  
 111266 117393: contig of 6127 bp in length  
 117394 117492: gap of 100 bp  
 117493 128939: contig of 11447 bp in length  
 128940 129039: gap of 100 bp  
 129040 137760: contig of 8721 bp in length  
 137761 137860: gap of 100 bp  
 137861 149949: contig of 12089 bp in length  
 149950 150049: gap of 100 bp  
 150050 164298: contig of 14249 bp in length  
 164299 164398: gap of 100 bp  
 164399 164490: contig of 92 bp in length.

FEATURES

source

1..16490  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="5"  
 /map="s"  
 /clone="RP11-424A10"  
 /clone\_lib="RPC1-11 Human Male BAC"

misc\_feature

1..379  
 /note="assembly\_fragment"  
 clone\_end:SP6  
 vector\_side:left

misc\_feature

480..15053  
 /note="assembly\_fragment"

misc\_feature

15154..15720  
 /note="assembly\_fragment"

misc\_feature

15821..17135  
 /note="assembly\_fragment"

misc\_feature

17236..18343  
 /note="assembly\_fragment"

misc\_feature

18444..19570  
 /note="assembly\_fragment"

Query Match 15.8%; Score 42.6; DB 2; Length 164490;

```

Best Local Similarity 48.5%; Pred. No. 0.71;
Matches 117; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY 2 GGAGCAATAAATCGTTATTAAAGATAATGGTATAGAACGAGCTTCGATGAATCAATGA 61
Db 32234 GTAGCAATTCATCAATTCATCAATGATGATGCTTACTACAGAGGAAAG 32293
QY 62 TTTTATTGAGATCGTGGGACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 121
Db 32294 CAATTCGTGCTCTATGCAAAATTTATATAAGTAACCAAAATGGCTACAGGTGGGTGA 32353
QY 122 AATTCACAGGGTAAAGAGGCTGCTGTAATTCGATTCATTCGGCTTGGCAAAAT 181
Db 32354 GAGTCCTACAGGAGATTCCTCTAGCTCAGTTGGTACGATGATGACTCAGAAATCAGG 32413
QY 182 TGAATTTCTACTATCCAGGGAACGAAAGGTTACTTTTACTTTGGATTATCGAGAT 241
Db 32414 AGGATTTAGTGTGTTCCAAAGAAACAGATTTTGTGTCATGTTTGAATATAGATATAT 32473
QY 242 T 242
Db 32474 T 32474

AC061968 63215 bp DNA linear HTG 21-APR-2000
Homo sapiens chromosome 9 clone RP11-167K13 map 9, LOW-PASS
SEQUENCE SAMPLING.
AC061968
VERSION
HTG: HTGS_PHASE0.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
1 (bases 1 to 63215)
Homo sapiens chromosome 9, clone RP11-167K13
Unpublished
2 (bases 1 to 63215)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Bouckgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collumore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodgson,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginder,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheters,R.,
Meltrim,J., Meneus,L., Mihova,P., Miranda,C., Mianga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigglio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (21-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

```

```

----- Project Information
Center project name: L7248
Center clone name: 167_K_13
-----
* NOTE: This record contains 80 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
* 1
* 710: contig of 710 bp in length
* 711: gap of 100 bp
* 811: contig of 691 bp in length
* 1501: gap of 100 bp
* 1502: gap of 100 bp
* 1503: gap of 100 bp
* 1504: gap of 100 bp
* 1505: gap of 100 bp
* 1506: gap of 100 bp
* 1507: gap of 100 bp
* 1508: gap of 100 bp
* 1509: gap of 100 bp
* 1510: gap of 100 bp
* 1511: gap of 100 bp
* 1512: gap of 100 bp
* 1513: gap of 100 bp
* 1514: gap of 100 bp
* 1515: gap of 100 bp
* 1516: gap of 100 bp
* 1517: gap of 100 bp
* 1518: gap of 100 bp
* 1519: gap of 100 bp
* 1520: gap of 100 bp
* 1521: gap of 100 bp
* 1522: gap of 100 bp
* 1523: gap of 100 bp
* 1524: gap of 100 bp
* 1525: gap of 100 bp
* 1526: gap of 100 bp
* 1527: gap of 100 bp
* 1528: gap of 100 bp
* 1529: gap of 100 bp
* 1530: gap of 100 bp
* 1531: gap of 100 bp
* 1532: gap of 100 bp
* 1533: gap of 100 bp
* 1534: gap of 100 bp
* 1535: gap of 100 bp
* 1536: gap of 100 bp
* 1537: gap of 100 bp
* 1538: gap of 100 bp
* 1539: gap of 100 bp
* 1540: gap of 100 bp
* 1541: gap of 100 bp
* 1542: gap of 100 bp
* 1543: gap of 100 bp
* 1544: gap of 100 bp
* 1545: gap of 100 bp
* 1546: gap of 100 bp
* 1547: gap of 100 bp
* 1548: gap of 100 bp
* 1549: gap of 100 bp
* 1550: gap of 100 bp
* 1551: gap of 100 bp
* 1552: gap of 100 bp
* 1553: gap of 100 bp
* 1554: gap of 100 bp
* 1555: gap of 100 bp
* 1556: gap of 100 bp
* 1557: gap of 100 bp
* 1558: gap of 100 bp
* 1559: gap of 100 bp
* 1560: gap of 100 bp
* 1561: gap of 100 bp
* 1562: gap of 100 bp
* 1563: gap of 100 bp
* 1564: gap of 100 bp
* 1565: gap of 100 bp
* 1566: gap of 100 bp
* 1567: gap of 100 bp
* 1568: gap of 100 bp
* 1569: gap of 100 bp
* 1570: gap of 100 bp
* 1571: gap of 100 bp
* 1572: gap of 100 bp
* 1573: gap of 100 bp
* 1574: gap of 100 bp
* 1575: gap of 100 bp
* 1576: gap of 100 bp
* 1577: gap of 100 bp
* 1578: gap of 100 bp
* 1579: gap of 100 bp
* 1580: gap of 100 bp
* 1581: gap of 100 bp
* 1582: gap of 100 bp
* 1583: gap of 100 bp
* 1584: gap of 100 bp
* 1585: gap of 100 bp
* 1586: gap of 100 bp
* 1587: gap of 100 bp
* 1588: gap of 100 bp
* 1589: gap of 100 bp
* 1590: gap of 100 bp
* 1591: gap of 100 bp
* 1592: gap of 100 bp
* 1593: gap of 100 bp
* 1594: gap of 100 bp
* 1595: gap of 100 bp
* 1596: gap of 100 bp
* 1597: gap of 100 bp
* 1598: gap of 100 bp
* 1599: gap of 100 bp
* 1600: gap of 100 bp
* 1601: gap of 100 bp
* 1602: gap of 100 bp
* 1603: gap of 100 bp
* 1604: gap of 100 bp
* 1605: gap of 100 bp
* 1606: gap of 100 bp
* 1607: gap of 100 bp
* 1608: gap of 100 bp
* 1609: gap of 100 bp
* 1610: gap of 100 bp
* 1611: gap of 100 bp
* 1612: gap of 100 bp
* 1613: gap of 100 bp
* 1614: gap of 100 bp
* 1615: gap of 100 bp
* 1616: gap of 100 bp
* 1617: gap of 100 bp
* 1618: gap of 100 bp
* 1619: gap of 100 bp
* 1620: gap of 100 bp
* 1621: gap of 100 bp
* 1622: gap of 100 bp
* 1623: gap of 100 bp
* 1624: gap of 100 bp
* 1625: gap of 100 bp
* 1626: gap of 100 bp
* 1627: gap of 100 bp
* 1628: gap of 100 bp
* 1629: gap of 100 bp
* 1630: gap of 100 bp
* 1631: gap of 100 bp
* 1632: gap of 100 bp
* 1633: gap of 100 bp
* 1634: gap of 100 bp
* 1635: gap of 100 bp
* 1636: gap of 100 bp
* 1637: gap of 100 bp
* 1638: gap of 100 bp
* 1639: gap of 100 bp
* 1640: gap of 100 bp
* 1641: gap of 100 bp
* 1642: gap of 100 bp
* 1643: gap of 100 bp
* 1644: gap of 100 bp
* 1645: gap of 100 bp
* 1646: gap of 100 bp
* 1647: gap of 100 bp
* 1648: gap of 100 bp
* 1649: gap of 100 bp
* 1650: gap of 100 bp
* 1651: gap of 100 bp
* 1652: gap of 100 bp
* 1653: gap of 100 bp
* 1654: gap of 100 bp
* 1655: gap of 100 bp
* 1656: gap of 100 bp
* 1657: gap of 100 bp
* 1658: gap of 100 bp
* 1659: gap of 100 bp
* 1660: gap of 100 bp
* 1661: gap of 100 bp
* 1662: gap of 100 bp
* 1663: gap of 100 bp
* 1664: gap of 100 bp
* 1665: gap of 100 bp
* 1666: gap of 100 bp
* 1667: gap of 100 bp
* 1668: gap of 100 bp
* 1669: gap of 100 bp
* 1670: gap of 100 bp
* 1671: gap of 100 bp
* 1672: gap of 100 bp
* 1673: gap of 100 bp
* 1674: gap of 100 bp
* 1675: gap of 100 bp
* 1676: gap of 100 bp
* 1677: gap of 100 bp
* 1678: gap of 100 bp
* 1679: gap of 100 bp
* 1680: gap of 100 bp
* 1681: gap of 100 bp
* 1682: gap of 100 bp
* 1683: gap of 100 bp
* 1684: gap of 100 bp
* 1685: gap of 100 bp
* 1686: gap of 100 bp
* 1687: gap of 100 bp
* 1688: gap of 100 bp
* 1689: gap of 100 bp
* 1690: gap of 100 bp
* 1691: gap of 100 bp
* 1692: gap of 100 bp
* 1693: gap of 100 bp
* 1694: gap of 100 bp
* 1695: gap of 100 bp
* 1696: gap of 100 bp
* 1697: gap of 100 bp
* 1698: gap of 100 bp
* 1699: gap of 100 bp
* 1700: gap of 100 bp
* 1701: gap of 100 bp
* 1702: gap of 100 bp
* 1703: gap of 100 bp
* 1704: gap of 100 bp
* 1705: gap of 100 bp
* 1706: gap of 100 bp
* 1707: gap of 100 bp
* 1708: gap of 100 bp
* 1709: gap of 100 bp
* 1710: gap of 100 bp
* 1711: gap of 100 bp
* 1712: gap of 100 bp
* 1713: gap of 100 bp
* 1714: gap of 100 bp
* 1715: gap of 100 bp
* 1716: gap of 100 bp
* 1717: gap of 100 bp
* 1718: gap of 100 bp
* 1719: gap of 100 bp
* 1720: gap of 100 bp
* 1721: gap of 100 bp
* 1722: gap of 100 bp
* 1723: gap of 100 bp
* 1724: gap of 100 bp
* 1725: gap of 100 bp
* 1726: gap of 100 bp
* 1727: gap of 100 bp
* 1728: gap of 100 bp
* 1729: gap of 100 bp
* 1730: gap of 100 bp
* 1731: gap of 100 bp
* 1732: gap of 100 bp
* 1733: gap of 100 bp
* 1734: gap of 100 bp
* 1735: gap of 100 bp
* 1736: gap of 100 bp
* 1737: gap of 100 bp
* 1738: gap of 100 bp
* 1739: gap of 100 bp
* 1740: gap of 100 bp
* 1741: gap of 100 bp
* 1742: gap of 100 bp
* 1743: gap of 100 bp
* 1744: gap of 100 bp
* 1745: gap of 100 bp
* 1746: gap of 100 bp
* 1747: gap of 100 bp
* 1748: gap of 100 bp
* 1749: gap of 100 bp
* 1750: gap of 100 bp
* 1751: gap of 100 bp
* 1752: gap of 100 bp
* 1753: gap of 100 bp
* 1754: gap of 100 bp
* 1755: gap of 100 bp
* 1756: gap of 100 bp
* 1757: gap of 100 bp
* 1758: gap of 100 bp
* 1759: gap of 100 bp
* 1760: gap of 100 bp
* 1761: gap of 100 bp
* 1762: gap of 100 bp
* 1763: gap of 100 bp
* 1764: gap of 100 bp
* 1765: gap of 100 bp
* 1766: gap of 100 bp
* 1767: gap of 100 bp
* 1768: gap of 100 bp
* 1769: gap of 100 bp
* 1770: gap of 100 bp
* 1771: gap of 100 bp
* 1772: gap of 100 bp
* 1773: gap of 100 bp
* 1774: gap of 100 bp
* 1775: gap of 100 bp
* 1776: gap of 100 bp
* 1777: gap of 100 bp
* 1778: gap of 100 bp
* 1779: gap of 100 bp
* 1780: gap of 100 bp
* 1781: gap of 100 bp
* 1782: gap of 100 bp
* 1783: gap of 100 bp
* 1784: gap of 100 bp
* 1785: gap of 100 bp
* 1786: gap of 100 bp
* 1787: gap of 100 bp
* 1788: gap of 100 bp
* 1789: gap of 100 bp
* 1790: gap of 100 bp
* 1791: gap of 100 bp
* 1792: gap of 100 bp
* 1793: gap of 100 bp
* 1794: gap of 100 bp
* 1795: gap of 100 bp
* 1796: gap of 100 bp
* 1797: gap of 100 bp
* 1798: gap of 100 bp
* 1799: gap of 100 bp
* 1800: gap of 100 bp
* 1801: gap of 100 bp
* 1802: gap of 100 bp
* 1803: gap of 100 bp
* 1804: gap of 100 bp
* 1805: gap of 100 bp
* 1806: gap of 100 bp
* 1807: gap of 100 bp
* 1808: gap of 100 bp
* 1809: gap of 100 bp
* 1810: gap of 100 bp
* 1811: gap of 100 bp
* 1812: gap of 100 bp
* 1813: gap of 100 bp
* 1814: gap of 100 bp
* 1815: gap of 100 bp
* 1816: gap of 100 bp
* 1817: gap of 100 bp
* 1818: gap of 100 bp
* 1819: gap of 100 bp
* 1820: gap of 100 bp
* 1821: gap of 100 bp
* 1822: gap of 100 bp
* 1823: gap of 100 bp
* 1824: gap of 100 bp
* 1825: gap of 100 bp
* 1826: gap of 100 bp
* 1827: gap of 100 bp
* 1828: gap of 100 bp
* 1829: gap of 100 bp
* 1830: gap of 100 bp
* 1831: gap of 100 bp
* 1832: gap of 100 bp
* 1833: gap of 100 bp
* 1834: gap of 100 bp
* 1835: gap of 100 bp
* 1836: gap of 100 bp
* 1837: gap of 100 bp
* 1838: gap of 100 bp
* 1839: gap of 100 bp
* 1840: gap of 100 bp
* 1841: gap of 100 bp
* 1842: gap of 100 bp
* 1843: gap of 100 bp
* 1844: gap of 100 bp
* 1845: gap of 100 bp
* 1846: gap of 100 bp
* 1847: gap of 100 bp
* 1848: gap of 100 bp
* 1849: gap of 100 bp
* 1850: gap of 100 bp
* 1851: gap of 100 bp
* 1852: gap of 100 bp
* 1853: gap of 100 bp
* 1854: gap of 100 bp
* 1855: gap of 100 bp
* 1856: gap of 100 bp
* 1857: gap of 100 bp
* 1858: gap of 100 bp
* 1859: gap of 100 bp
* 1860: gap of 100 bp
* 1861: gap of 100 bp
* 1862: gap of 100 bp
* 1863: gap of 100 bp
* 1864: gap of 100 bp
* 1865: gap of 100 bp
* 1866: gap of 100 bp
* 1867: gap of 100 bp
* 1868: gap of 100 bp
* 1869: gap of 100 bp
* 1870: gap of 100 bp
* 1871: gap of 100 bp
* 1872: gap of 100 bp
* 1873: gap of 100 bp
* 1874: gap of 100 bp
* 1875: gap of 100 bp
* 1876: gap of 100 bp
* 1877: gap of 100 bp
* 1878: gap of 100 bp
* 1879: gap of 100 bp
* 1880: gap of 100 bp
* 1881: gap of 100 bp
* 1882: gap of 100 bp
* 1883: gap of 100 bp
* 1884: gap of 100 bp
* 1885: gap of 100 bp
* 1886: gap of 100 bp
* 1887: gap of 100 bp
* 1888: gap of 100 bp
* 1889: gap of 100 bp
* 1890: gap of 100 bp
* 1891: gap of 100 bp
* 1892: gap of 100 bp
* 1893: gap of 100 bp
* 1894: gap of 100 bp
* 1895: gap of 100 bp
* 1896: gap of 100 bp
* 1897: gap of 100 bp
* 1898: gap of 100 bp
* 1899: gap of 100 bp
* 1900: gap of 100 bp
* 1901: gap of 100 bp
* 1902: gap of 100 bp
* 1903: gap of 100 bp
* 1904: gap of 100 bp
* 1905: gap of 100 bp
* 1906: gap of 100 bp
* 1907: gap of 100 bp
* 1908: gap of 100 bp
* 1909: gap of 100 bp
* 1910: gap of 100 bp
* 1911: gap of 100 bp
* 1912: gap of 100 bp
* 1913: gap of 100 bp
* 1914: gap of 100 bp
* 1915: gap of 100 bp
* 1916: gap of 100 bp
* 1917: gap of 100 bp
* 1918: gap of 100 bp
* 1919: gap of 100 bp
* 1920: gap of 100 bp
* 1921: gap of 100 bp
* 1922: gap of 100 bp
* 1923: gap of 100 bp
* 1924: gap of 100 bp
* 1925: gap of 100 bp
* 1926: gap of 100 bp
* 1927: gap of 100 bp
* 1928: gap of 100 bp
* 1929: gap of 100 bp
* 1930: gap of 100 bp
* 1931: gap of 100 bp
* 1932: gap of 100 bp
* 1933: gap of 100 bp
* 1934: gap of 100 bp
* 1935: gap of 100 bp
* 1936: gap of 100 bp
* 1937: gap of 100 bp
* 1938: gap of 100 bp
* 1939: gap of 100 bp
* 1940: gap of 100 bp
* 1941: gap of 100 bp
* 1942: gap of 100 bp
* 1943: gap of 100 bp
* 1944: gap of 100 bp
* 1945: gap of 100 bp
* 1946: gap of 100 bp
* 1947: gap of 100 bp
* 1948: gap of 100 bp
* 1949: gap of 100 bp
* 1950: gap of 100 bp
* 1951: gap of 100 bp
* 1952: gap of 100 bp
* 1953: gap of 100 bp
* 1954: gap of 100 bp
* 1955: gap of 100 bp
* 1956: gap of 100 bp
* 1957: gap of 100 bp
* 1958: gap of 100 bp
* 1959: gap of 100 bp
* 1960: gap of 100 bp
* 1961: gap of 100 bp
* 1962: gap of 100 bp
* 1963: gap of 100 bp
* 1964: gap of 100 bp
* 1965: gap of 100 bp
* 1966: gap of 100 bp
* 1967: gap of 100 bp
* 1968: gap of 100 bp
* 1969: gap of 100 bp
* 1970: gap of 100 bp
* 1971: gap of 100 bp
* 1972: gap of 100 bp
* 1973: gap of 100 bp
* 1974: gap of 100 bp
* 1975: gap of 100 bp
* 1976: gap of 100 bp
* 1977: gap of 100 bp
* 1978: gap of 100 bp
* 1979: gap of 100 bp
* 1980: gap of 100 bp
* 1981: gap of 100 bp
* 1982: gap of 100 bp
* 1983: gap of 100 bp
* 1984: gap of 100 bp
* 1985: gap of 100 bp
* 1986: gap of 100 bp
* 1987: gap of 100 bp
* 1988: gap of 100 bp
* 1989: gap of 100 bp
* 1990: gap of 100 bp
* 1991: gap of 100 bp
* 1992: gap of 100 bp
* 1993: gap of 100 bp
* 1994: gap of 100 bp
* 1995: gap of 100 bp
* 1996: gap of 100 bp
* 1997: gap of 100 bp
* 1998: gap of 100 bp
* 1999: gap of 100 bp
* 2000: gap of 100 bp
* 2001: gap of 100 bp
* 2002: gap of 100 bp
* 2003: gap of 100 bp
* 2004: gap of 100 bp
* 2005: gap of 100 bp
* 2006: gap of 100 bp
* 2007: gap of 100 bp
* 2008: gap of 100 bp
* 2009: gap of 100 bp
* 2010: gap of 100 bp
* 2011: gap of 100 bp
* 2012: gap of 100 bp
* 2013: gap of 100 bp
* 2014: gap of 100 bp
* 2015: gap of 100 bp
* 2016: gap of 100 bp
* 2017: gap of 100 bp
* 2018: gap of 100 bp
* 2019: gap of 100 bp
* 2020: gap of 100 bp
* 2021: gap of 100 bp
* 2022: gap of 100 bp
* 2023: gap of 100 bp
* 2024: gap of 100 bp
* 2025: gap of 100 bp
* 2026: gap of 100 bp
* 2027: gap of 100 bp
* 2028: gap of 100 bp
* 2029: gap of 100 bp
* 2030: gap of 100 bp
* 2031: gap of 100 bp
* 2032: gap of 100 bp
* 2033: gap of 100 bp
* 2034: gap of 100 bp
* 2035: gap of 100 bp
* 2036: gap of 100 bp
* 2037: gap of 100 bp
* 2038: gap of 100 bp
* 2039: gap of 100 bp
* 2040: gap of 100 bp
* 2041: gap of 100 bp
* 2042: gap of 100 bp
* 2043: gap of 100 bp
* 2044: gap of 100 bp
* 2045: gap of 100 bp
* 2046: gap of 100 bp
* 2047: gap of 100 bp
* 2048: gap of 100 bp
* 2049: gap of 100 bp
* 2050: gap of 100 bp
* 2051: gap of 100 bp
* 2052: gap of 100 bp
* 2053: gap of 100 bp
* 2054: gap of 100 bp
* 2055: gap of 100 bp
* 2056: gap of 100 bp
* 2057: gap of 100 bp
* 2058: gap of 100 bp
* 2059: gap of 100 bp
* 2060: gap of 100 bp
* 2061: gap of 100 bp
* 2062: gap of 100 bp
* 2063: gap of 100 bp
* 2064: gap of 100 bp
* 2065: gap of 100 bp
* 2066: gap of 100 bp
* 2067: gap of 100 bp
* 2068: gap of 100 bp
* 2069: gap of 100 bp
* 2070: gap of 100 bp
* 2071: gap of 100 bp
* 2072: gap of 100 bp
* 2073: gap of 100 bp
* 2074: gap of 100 bp
* 2075: gap of 100 bp
* 2076: gap of 100 bp
* 2077: gap of 100 bp
* 2078: gap of 100 bp
* 2079: gap of 100 bp
* 2080: gap of 100 bp
* 2081: gap of 100 bp
* 2082: gap of 100 bp
* 2083: gap of 100 bp
* 2084: gap of 100 bp
* 2085: gap of 100 bp
* 2086: gap of 100 bp
* 2087: gap of 100 bp
* 2088: gap of 100 bp
* 2089: gap of 100 bp
* 2090: gap of 100 bp
* 2091: gap of 100 bp
* 2092: gap of 100 bp
* 2093: gap of 100 bp
* 2094: gap of 100 bp
* 2095: gap of 100 bp
* 2096: gap of 100 bp
* 2097: gap of 100 bp
* 2098: gap of 100 bp
* 2099: gap of 100 bp
* 2100: gap of 100 bp
* 2101: gap of 100 bp
* 2102: gap of 100 bp
* 2103: gap of 100 bp
* 2104: gap of 100 bp
* 2105: gap of 100 bp
* 2106: gap of 100 bp
* 2107: gap of 100 bp
* 2108: gap of 100 bp
* 2109: gap of 100 bp
* 2110: gap of 100 bp
* 2111: gap of 100 bp
* 2112: gap of 100 bp
* 2113: gap of 100 bp
* 2114: gap of 100 bp
* 2115: gap of 100 bp
* 2116: gap of 100 bp
* 2117: gap of 100 bp
* 2118: gap of 100 bp
* 2119: gap of 100 bp
* 2120: gap of 100 bp
* 2121: gap of 100 bp
* 2122: gap of 100 bp
* 2123: gap of 100 bp
* 2124: gap of 100 bp
* 2125: gap of 100 bp
* 2126: gap of 100 bp
* 2127: gap of 100 bp
* 2128: gap of 100 bp
* 2129: gap of 100 bp
* 2130: gap of 100 bp
* 2131: gap of 100 bp
* 2132: gap of 100 bp
* 2133: gap of 100 bp
* 2134: gap of 100 bp
* 2135: gap of 100 bp
* 2136: gap of 100 bp
* 2137: gap of 100 bp
* 2138: gap of 100 bp
* 2139: gap of 100 bp
* 2140: gap of 100 bp
* 2141: gap of 100 bp
* 2142: gap of 100 bp
* 2143: gap of 100 bp
* 2144: gap of 100 bp
* 2145: gap of 100 bp
* 2146: gap of 100 bp
* 2147: gap of 100 bp
* 2148: gap of 100 bp
* 2149: gap of 100 bp
* 2150: gap of 100 bp
* 2151: gap of 100 bp
* 2152: gap of 100 bp
* 2153: gap of 100 bp
* 2154: gap of 100 bp
* 2155: gap of 100 bp
* 2156: gap of 100 bp
* 2157: gap of 100 bp
* 2158: gap of 100 bp
* 2159: gap of 100 bp
* 2160: gap of 100 bp
* 2161: gap of 100 bp
* 2162: gap of 100 bp
* 2163: gap of 100 bp
* 2164: gap of 100 bp
* 2165: gap of 100 bp
* 2166: gap of 100 bp
* 2167: gap of 100 bp
* 2168: gap of 100 bp
* 2169: gap of 100 bp
* 2170: gap of 100 bp
* 2171: gap of 100 bp
* 2172: gap of 100 bp
* 2173: gap of 100 bp
* 2174: gap of 100 bp
* 2175: gap of 100 bp
* 2176: gap of 100 bp
* 2177: gap of 100 bp
* 2178: gap of 100 bp
* 2179: gap of 100 bp
* 2180: gap of 100 bp
* 2181: gap of 100 bp
* 2182: gap of 100 bp
* 2183: gap of 100 bp
* 2184: gap of 100 bp
* 2185: gap of 100 bp
* 2186: gap of 100 bp
* 2187: gap of 100 bp
* 2188: gap of 100 bp
* 2189: gap of 100 bp
* 2190: gap of 100 bp
* 2191: gap of 100 bp
* 2192: gap of 100 bp
* 2193: gap of 100 bp
* 2194: gap of 100 bp
* 2195: gap of 100 bp
* 2196: gap of 100 bp
* 2197: gap of 100 bp
* 2198: gap of 100 bp
* 2199: gap of 100 bp
* 2200: gap of 100 bp
* 2201: gap of 100 bp
* 2202: gap of 100 bp
* 2203: gap of 100 bp
* 2204: gap of 100 bp
* 2205: gap of 100 bp
* 2206: gap of 100 bp
* 2207: gap of 100 bp
* 2208: gap of 100 bp
* 2209: gap of 100 bp
* 2210: gap of 100 bp
* 2211: gap of 100 bp
* 2212: gap of 100 bp
* 2213: gap of 100 bp
* 2214: gap of 100 bp
* 2215: gap of 100 bp
* 2216: gap of 100 bp
* 2217
```

```

* 22959 23637: contig of 679 bp in length
* 23638 23737: gap of 100 bp
* 23738 24409: contig of 672 bp in length
* 24410 24509: gap of 100 bp
* 24510 25197: contig of 688 bp in length
* 25198 25297: gap of 100 bp
* 25298 26006: contig of 709 bp in length
* 26007 26106: gap of 100 bp
* 26107 26801: contig of 695 bp in length
* 26802 26901: gap of 100 bp
* 26902 27593: contig of 692 bp in length
* 27594 27693: gap of 100 bp
* 27694 28380: contig of 687 bp in length
* 28381 28480: gap of 100 bp
* 28481 29176: contig of 696 bp in length
* 29177 29276: gap of 100 bp
* 29277 29988: contig of 712 bp in length
* 29989 30088: gap of 100 bp
* 30089 30776: contig of 688 bp in length
* 30777 30876: gap of 100 bp
* 30877 31588: contig of 682 bp in length
* 31589 31688: gap of 100 bp
* 31689 32325: contig of 667 bp in length
* 32326 32425: gap of 100 bp
* 32426 33122: contig of 697 bp in length
* 33123 33222: gap of 100 bp
* 33223 33870: contig of 748 bp in length
* 33871 34070: gap of 100 bp
* 34071 34772: contig of 702 bp in length
* 34773 34872: gap of 100 bp
* 34873 35574: contig of 702 bp in length
* 35575 35674: gap of 100 bp
* 35675 36358: contig of 684 bp in length
* 36359 36458: gap of 100 bp
* 36459 37137: contig of 679 bp in length
* 37138 37237: gap of 100 bp
* 37238 37936: contig of 699 bp in length
* 37937 38036: gap of 100 bp
* 38037 38742: contig of 706 bp in length
* 38743 38842: gap of 100 bp
* 38843 39536: contig of 694 bp in length
* 39537 39636: gap of 100 bp
* 39637 40294: contig of 658 bp in length
* 40295 40394: gap of 100 bp
* 40395 41073: contig of 679 bp in length
* 41074 41173: gap of 100 bp
* 41174 41868: contig of 695 bp in length
* 41869 41968: gap of 100 bp
* 41969 42677: contig of 709 bp in length
* 42678 42777: gap of 100 bp
* 42778 43476: contig of 699 bp in length
* 43477 43576: gap of 100 bp
* 43577 44265: contig of 689 bp in length
* 44266 44365: gap of 100 bp
* 44366 45064: contig of 699 bp in length
* 45065 45164: gap of 100 bp
* 45165 45846: contig of 682 bp in length
* 45847 45946: gap of 100 bp
* 45947 46652: contig of 706 bp in length
* 46653 46752: gap of 100 bp
* 46753 47435: contig of 683 bp in length
* 47436 47535: gap of 100 bp
* 47536 48208: contig of 673 bp in length
* 48209 48308: gap of 100 bp
* 48309 48983: contig of 675 bp in length
* 48984 49083: gap of 100 bp
* 49084 49784: contig of 701 bp in length
* 49785 49884: gap of 100 bp
* 49885 50623: contig of 741 bp in length
* 50624 50723: gap of 100 bp
* 50725 51418: contig of 693 bp in length
* 51419 51519: gap of 100 bp
* 51519 52199: contig of 681 bp in length

```

```

* 52200 52299: gap of 100 bp
* 52300 52982: contig of 683 bp in length
* 52983 53082: gap of 100 bp
* 53083 53800: contig of 718 bp in length
* 53801 53900: gap of 100 bp
* 53901 54619: contig of 719 bp in length
* 54620 54719: gap of 100 bp

Query Match      15.7%; Score 42.2; DB 2; Length 63215;
Best Local Similarity 48.1%; Pred. No. 0.98;
Matches 116; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

QY 2 GGAGCATAAATCGTTATTAAAGATAATGGTATAGGAACGAGCTTCGATCAAAATCAATCA 61
Db 15376 GTAGCATTTTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 15435

QY 62 TTTTATTGAGAAATCGTCCGGAACAGAGGGAAGAAACAGCTCCCTCGGGAAG 121
Db 15436 CAATTGGTGCCTATGCAAAATTTATATAAGTAACCAAAATGGCTACAGTGGGTGA 15495

QY 122 AATCCAAAGGTAATAAAGGCGCTTCGTAATTTGGCAATTTCCGGCTTGGCAACAAAT 181
Db 15496 GAGTCTTACAGGAAGATTCTTCTACTCAGTTGGTACAGTATGCTACANAATCAGG 15555

QY 182 TGAATTTCTACTATCCAGGAAACGAAAGGTTACTTTTACTTTGGATTATCGAGAT 241
Db 15556 AGGGATTAGTGTTCCTCCAAAGAAACAGATTTTGTCTATGCTTTAGATATAGATATAT 15615

QY 242 T 242
Db 15616 T 15616

```

## RESULT 14

```

AC018349/c
LOCUS      AC018349      185539 bp      DNA      linear      HTG 25-MAR-2001
DEFINITION Homo sapiens chromosome 5 clone RP11-140N7, WORKING DRAFT SEQUENCE,
4 unordered pieces.
ACCESSION  AC018349
VERSION     AC018349.15  GI:13446260
KEYWORDS   HTG; HTGS PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 185539)
AUTHORS     Abola, A.P., Bruno, D., Conn, L., Della Rosa, M., Faulkner, D.,
Federspiel, N., Glukhov, S., Hansen, N., Herman, Z.S., Hyman, R.,
Mao, J., Komp, C., Kottler, S., Lam, B., Marathe, R., Miranda, M.,
Morehouse, A.J., Nguyen, M., Oefner, P., Palm, C.J., Ramirez, D.,
Southwick, A.M., Webb, C., Wilhelmy, J., Yu, S. and Davis, R.W.
Unpublished
JOURNAL     2 (bases 1 to 185539)
REFERENCE    Bruno, D., Conn, L., Della Rosa, M., Federspiel, N., Foreman, P.,
Glukhov, S., Hansen, N., Hyman, R., Mao, J., Marathe, R.,
Morehouse, A.J., Oefner, P., Palm, C.J., Ramirez, D., Wilhelmy, J.,
Yu, S. and Davis, R.W.
Direct Submission
TITLE       Submitted (09-DEC-1999) DNA Sequencing and Technology Center,
JOURNAL     Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA
COMMENT     On Mar 25, 2001 this sequence version replaced gi:13270572.
-----
Center: Genome Center
Center: Stanford DNA Sequencing and Technology Development
Center
Center code: SDSTDC
Web site: http://sequence-www.stanford.edu/group/human/
Contact: hum-info@sequence.stanford.edu
-----
Project Information
Center project name: 736
Center clone name: RP11-140N7
-----
Summary Statistics
Sequencing Vector: M13mp18; X02513; 99% of reads

```



```

Sequencing Vector: plasmid; plasmid_accession; 1% of reads
Chemistry: Dye-primer; 2% of reads
Assembly: Dye-terminator BigDye; 97% of reads
Consensus quality: 182658 bases at least Q40
Consensus quality: 183349 bases at least Q30
Consensus quality: 183837 bases at least Q20
Insert size: 17966; agarose-fp
Insert size: 185239; sum-of-contigs
Quality coverage: 9.2x in Q20 bases; agarose-fp
Quality coverage: 8.8x in Q20 bases; sum-of-contigs.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
1 3543: contig of 3543 bp in length
3544 3643: gap of unknown length
3644 39213: contig of 35570 bp in length
39214 39313: gap of unknown length
39314 87111: contig of 47798 bp in length
87112 87211: gap of unknown length
87212 185539: contig of 98328 bp in length.
Location/Qualifiers
1..185539
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-140N7"
/clone_lib="RPC1 human BAC library 11"
1..3543
/note="assembly_name:Contig12
clone_end:77"
3644..39213
/note="assembly_name:Contig13
clone_end:SP6"
39314..87111
/note="assembly_name:Contig14"
87212..185539
/note="assembly_name:Contig15"
BASE COUNT 55079 a 38207 c 39983 g 52966 t 304 others
ORIGIN
Query Match 15.2%; Score 41; DB 2; Length 185539;
Best Local Similarity 48.1%; Pred. No. 1.9;
Matches 116; Conservative 0; Mismatches 125; Indels 0; Gaps 0;
QY 2 GGAGCATAAATCGTTATTAAAGATAATGGTATAGGAACGAGCTTCGATGAATCAATGA 61
Db 73259 GTAGCAATTTGATCATTCCTCAATGATACGATTCGATGGCTTACTACAGAGGAAAG 73200
QY 62 TTTTATTGAGATCGGTGCGACAGAGGAGGAGAAACAAAGCCTCCCGTCGCGAAG 121
Db 73199 CAATTGGTCTCTATGCAAAATTTATTAAGTAATACCAAGAAATGGCTACAGTGGGTGA 73140
QY 122 AATTCACCGGTAAAAAGGCTTGGTAAATTCGCATTATTCGGCTTGGCAACAAAAT 181
Db 73139 GAGTCCTACAGGAAGATTCCTCAGTCAGTCAGTGGTACAGTATGACTCAGAAATCAGG 73080
QY 182 TGAATTTCTACTATCCAGGAAACGAAAGGTTACTTTTACTTTGGATTATGCAGAGAT 241
Db 73079 AGGGATTTAGTGTTCCTCAAGAAACAGAGATTTCATGCTTTAGAAATATAGATAT 73020
QY 242 T 242
Db 73019 T 73019

```

RESULT 15

```

AC010279
LOCUS 125746 bp DNA linear PSI 23-FEB-2002
DEFINITION Homo sapiens chromosome 5 clone CTC-533D18, complete sequence.
AC010279
VERSION AC010279.5 GI:18873821
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 125746)
DOE Joint Genome Institute and Stanford Human Genome Center.
Direct Submission
JOURNAL Unpublished
2 (bases 1 to 125746)
DOE Joint Genome Institute.
Direct Submission
JOURNAL Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 125746)
DOE Joint Genome Institute and Stanford Human Genome Center.
Direct Submission
JOURNAL Submitted (30-MAR-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
4 (bases 1 to 125746)
DOE Joint Genome Institute and Stanford Human Genome Center.
Direct Submission
JOURNAL Submitted (23-FEB-2002) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
On Feb 23, 2002 this sequence version replaced gi:13489146.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >40 99.9% of Sequence;
Estimated Total Number of Errors is 0.1.
FEATURES
source
1..125746
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTC-533D18"
BASE COUNT 36432 a 25088 c 26461 g 37765 t
ORIGIN
Query Match 14.6%; Score 39.4; DB 9; Length 125746;
Best Local Similarity 47.7%; Pred. No. 5;
Matches 115; Conservative 0; Mismatches 126; Indels 0; Gaps 0;
QY 2 GGAGCATAAATCGTTATTAAAGATAATGGTATAGGAACGAGCTTCGATGAATCAATGA 61
Db 96193 GTAGCAATTTGATCATTCCTCAATGATACGATTCGATGGCTTACTACAGAGGAAAG 96252
QY 62 TTTTATTGAGATCGGTGCGAAGGAGGAAACAAAGCCTCCCGTCGCGAAG 121
Db 96253 CAATTGGTCTCTATGCAAAATTTATTAAGTAATACCAAGAAATGGCTACAGTGGGTGA 96312
QY 122 AATTCACCGGTAAAAAGGCTTGGTAAATTCGCATTATTCGGCTTGGCAACAAAAT 181
Db 96313 GAGTCCTACAGGAAGATTCCTCAGTCAGTCAGTGGTACAGTATGACTCAGAAATCAGG 96372
QY 182 TGAATTTCTACTATCCAGGAAACGAAAGGTTACTTTTACTTTGGATTATGCAGAGAT 241
Db 96373 AGGGATTTAGTGTTCCTCAAGAAACAGAGATTTCATGCTTTAGAAATATAGATAT 96432
QY 242 T 242
Db 96433 T 96433

```

Search completed: November 15, 2003, 03:33:52  
Job time : 1194.78 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 14, 2003, 23:55:19 ; Search time 173.613 Seconds  
(without alignments)  
4182.570 Million cell updates/sec

Title: US-09-928-457-77

Perfect score: 269

Sequence: 1 CGGAGCATAAATCGTTATT.....GCAAGGTATTATCAACG 269

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq 19Jun03.\*

```

1: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*
25: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	269	100.0	269	AAV03585	Neisseria meningitidis genomic fragment o
2	269	100.0	269	AAAI5389	DNA encoding a pol
3	264.2	98.2	858	AAAI5296	N. gonorrhoeae nuc
4	246.6	91.7	855	ABZ37755	DNA encoding a pol
5	246.6	91.7	858	AAAI5322	N. gonorrhoeae nuc
6	246.6	91.7	891	ABZ41353	Neisseria meningitidis genomic fragment o
7	121.8	45.3	270	AAV03596	
8	121.8	45.3	270	AAAI5393	

C 9	41	15.2	13443	22	ABA19170	Human nervous syst
C 10	41	15.2	13443	22	AAK81300	Human immune/haema
C 11	39.4	14.6	13451	22	ABA19169	Human nervous syst
C 12	39.4	14.6	13451	22	AAK81299	Human immune/haema
C 13	36.8	13.7	4951	23	ABL13122	Drosophila melanog
C 14	36.8	13.7	5395	23	ABL15458	Drosophila melanog
C 15	36	13.4	2000	24	ABZ16977	Arabidopsis thalia
C 16	36	13.4	185695	25	ABQ77405	Human THBS1 DNA
C 17	35.4	13.2	2944528	24	ABA03041	Listeria monocytog
C 18	34.8	12.9	1855	24	ABQ70544	Listeria monocytog
C 19	34.8	12.9	1855	24	ABQ71183	Listeria monocytog
C 20	34	12.6	381	22	AAF65964	Novel human polynu
C 21	33.8	12.6	733	22	AAK62701	Human immune/haema
C 22	33.6	12.5	7160	20	AAK86191	Nucleotide sequenc
C 23	33	12.3	1142	24	ABN89001	Human prostate spe
C 24	33	12.3	2729	22	AAK81442	Human immune/haema
C 25	32.8	12.2	29871	24	ABN86359	L. lactis multidru
C 26	32.6	12.1	1881	24	ABQ70528	Listeria monocytog
C 27	32.4	12.0	1664976	19	AAV21209	Methanococcus jann
C 28	32	11.9	1586	21	AAK36395	Arabidopsis thalia
C 29	32	11.9	6243	20	AAZ09598	Clostridium specie
C 30	32	11.9	30179	24	ABK47964	Human alcohol dehy
C 31	31.8	11.8	435	24	ABN90660	Staphylococcus epi
C 32	31.8	11.8	1355	25	ABZ37481	M13-III nucleotide
C 33	31.8	11.8	4583	25	ABV75375	Arabidopsis TES1 m
C 34	31.8	11.8	4590	25	ABV75374	Arabidopsis wild-t
C 35	31.8	11.8	6118	24	ABL33031	Human immune syste
C 36	31.8	11.8	13138	22	AAK65690	Human immune/haema
C 37	31.6	11.7	5033	25	ABZ09999	Haematopoietic cel
C 38	31.6	11.7	7847	23	ABL17450	Drosophila melanog
C 39	31.6	11.7	8033	22	AAK46307	Tumour suppressor
C 40	31.6	11.7	8033	24	ABL70169	Chemically treated
C 41	31.6	11.7	8033	24	AAK61116	Human gene regulat
C 42	31.6	11.7	8033	24	ABK31202	Signal transductio
C 43	31.4	11.7	281	24	ABQ55863	Human ovarian anti
C 44	31.4	11.7	581	21	AAK59135	Human secreted pro
C 45	31.4	11.7	884	22	AAH33290	Human colon cancer

#### ALIGNMENTS

```

RESULT 1
AAV03585
ID  AAV03585 standard; DNA; 269 BP.
XX
AC  AAV03585;
XX
DT  22-OCT-1998 (first entry)
XX
DE  Neisseria meningitidis DNA sequence C40.
XX
KW  N. gonorrhoeae; N. lactamica; chromosome Z2491; region 1; region 2;
KW  region 3; pathogenicity; blood-brain barrier; diagnosis; infection;
KW  meningitis; ss.
XX
OS  Neisseria meningitidis.
XX
PN  WO9802547-A2.
XX
PD  22-JAN-1998.
XX
PF  11-JUL-1997; 97WO-FR01295.
XX
PR  12-JUL-1996; 96FR-0008768.
XX
(PINRM ) INSERM INST NAT SANTE & RECH MEDICALE.
(PAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
(SWIK ) SMITHKLINE BEECHAM.
XX
PI  Nassif X, Tinsley C, Achtman M, Merker P, Ruelle J;
PI  Vinals C;
XX

```

DR WPI; 1998-110594/10..

XX Genes present in *Neisseria meningitidis* but not other *Neisseria*

PT species - and related host cells, RNA, anti-sense sequences,

PT polypeptide(s) and antibodies, useful for diagnosing *Neisseria*

PT meningitidis infection and in protective vaccines

XX

PS Example 4; Page 118; 150pp; French.

XX

CC AA03575-606 represent sequences that are present in *Neisseria*

CC meningitidis and *N. gonorrhoeae* but not in *N. lactamica*, except for the

CC genes involved in biosynthesis of the capsule polysaccharide, rPA or C,

CC opC, porA, rotamase, sequence IC106, IgA protease, pilin, pilC,

CC proteins which bind transferrin and opacity proteins. The DNA sequences

CC are responsible for the differences in pathogenicity between *N.*

CC meningitidis and *N. gonorrhoeae*, specifically they include the genes that

CC allow *N. meningitidis* to cross the blood-brain barrier. DNA sequences

CC common to *N. meningitidis* and *N. gonorrhoeae*, but absent from *N.*

CC lactamica, are responsible for colonisation and penetration of the

CC mucosa. The DNA sequences can be used to produce probes and primers, and

CC antibodies produced against the encoded proteins are used in standard

CC hybridisation/immunoassay processes for diagnosis of *N. meningitidis*

CC infection, particularly meningitis.

XX

SQ Sequence 269 BP; 95 A; 41 C; 65 G; 68 T; 0 other;

Query Match 100.0%; Score 269; DB 19; Length 269;

Best Local Similarity 100.0%; Pred. No. 5.5e-71;

Matches 269; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGAGCATAAATCGTTATTAAAGATATGTTATAGGAACGAGCTTCGATCAATCAATG 60

DB 1 CGGAGCATAAATCGTTATTAAAGATATGTTATAGGAACGAGCTTCGATCAATCAATG 60

QY 61 ATTTTATTTCGAGATCGTTCGGAACAGAGGGAAGAAACAAAGCTCCCGTGGGAA 120

DB 61 ATTTTATTTCGAGATCGTTCGGAACAGAGGGAAGAAACAAAGCTCCCGTGGGAA 120

QY 121 GAATTCGAACGGGTAAAGAGCCCTTGGTAAATGGCATTTATTCGGGCTTGGCAACAAA 180

DB 121 GAATTCGAACGGGTAAAGAGCCCTTGGTAAATGGCATTTATTCGGGCTTGGCAACAAA 180

QY 181 TTGAATTTCTACTATCCAGGGAACGAAGGGTTACTTTTACTTTGGATTATGCAGAGA 240

DB 181 TTGAATTTCTACTATCCAGGGAACGAAGGGTTACTTTTACTTTGGATTATGCAGAGA 240

QY 241 TTCGAAGAACGAGGTATTATCAACCG 269

DB 241 TTCGAAGAACGAGGTATTATCAACCG 269

RESULT 2

AAA15389

ID AAA15389 standard; DNA; 269 BP.

XX AAA15389;

AC

XX

DT 04-SEP-2000 (first entry)

XX

DE Genomic fragment of *Neisseria meningitidis* Z2491.

XX

KW Pathogenic strain; *Neisseria*; vaccine; *Neisseria* infection; ss.

XX

OS *Neisseria meningitidis*.

XX

PN WO200026375-A2.

XX

PD 11-MAY-2000.

XX

PF 28-OCT-1999; 99WO-FR02643.

XX

PR 30-OCT-1998; 98FR-0013693.

XX

PA (INMR) PASTEUR MERIEUX SERUMS & VACCINS SA.

PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.

XX

PI Aujame L, Bouchardon A, Renauld-Mongenie G, Rokbi B, Nassif X;

PI Tinsley C, Perrin A;

XX

DR WPI; 2000-365622/31.

XX

PT New polypeptide specific for pathogenic *Neisseria* useful in therapeutic

PT or preventative vaccines and for diagnosis

XX

XX Claim 1; Page 40; 187pp; French.

XX

CC The present sequence represents a genomic fragment of *Neisseria*

CC meningitidis. The specification describes proteins that are specific

CC for pathogenic strains of *Neisseria*. The polynucleotides, polypeptides,

CC or their antigenic fragments, are used in vaccines to treat or protect

CC against *Neisseria* infections, particularly by *N. meningitidis*. The

CC polynucleotide sequences are also used for recombinant production of

CC the polypeptide and to produce attenuated *Neisseria* strains that

CC overexpress it, or express it in a non-toxic mutant form.

CC note: the present sequence is included in Claim 1, but it is those

CC sequences that do not include the present sequence that are actually

CC claimed.

XX

SQ Sequence 269 BP; 95 A; 41 C; 65 G; 68 T; 0 other;

Query Match 100.0%; Score 269; DB 21; Length 269;

Best Local Similarity 100.0%; Pred. No. 5.5e-71;

Matches 269; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGAGCATAAATCGTTATTAAAGATATGTTATAGGAACGAGCTTCGATCAATCAATG 60

DB 1 CGGAGCATAAATCGTTATTAAAGATATGTTATAGGAACGAGCTTCGATCAATCAATG 60

QY 61 ATTTTATTTCGAGATCGTTCGGAACAGAGGGAAGAAACAAAGCTCCCGTGGGAA 120

DB 61 ATTTTATTTCGAGATCGTTCGGAACAGAGGGAAGAAACAAAGCTCCCGTGGGAA 120

QY 121 GAATTCGAACGGGTAAAGAGCCCTTGGTAAATGGCATTTATTCGGGCTTGGCAACAAA 180

DB 121 GAATTCGAACGGGTAAAGAGCCCTTGGTAAATGGCATTTATTCGGGCTTGGCAACAAA 180

QY 181 TTGAATTTCTACTATCCAGGGAACGAAGGGTTACTTTTACTTTGGATTATGCAGAGA 240

DB 181 TTGAATTTCTACTATCCAGGGAACGAAGGGTTACTTTTACTTTGGATTATGCAGAGA 240

QY 241 TTCGAAGAACGAGGTATTATCAACCG 269

DB 241 TTCGAAGAACGAGGTATTATCAACCG 269

RESULT 3

AAA15296

ID AAA15296 standard; DNA; 858 BP.

XX

AC AAA15296;

XX

DT 04-SEP-2000 (first entry)

XX

DE DNA encoding a polypeptide of a *Neisseria* pathogenic strain.

XX

KW Pathogenic strain; *Neisseria*; vaccine; *Neisseria* infection; ss.

XX

OS *Neisseria meningitidis*.

XX

PN Key Location/Qualifiers

XX CDS 1..858

XX /tag= a

XX

PN WO200026375-A2.

XX

PD 11-MAY-2000.

XX

```

XX 28-OCT-1999; 95WO-FR02643.
XX 30-OCT-1998; 98FR-0013693.
XX (INMR ) PASTEUR MERIEUX SERUMS & VACCINS SA.
XX (INMR ) INSERM INST NAT SANTE & RECH MEDICALE.
XX Autame L, Bouchardon A, Renauld-Mongenie G, Rokbi B, Naessif X,
XX Tinsley C, Perrin A;
XX WPI; 2000-365622/31.
XX P-PSDB; AAY93265.
XX New polypeptide specific for pathogenic Neisseria useful in therapeutic
XX or preventative vaccines and for diagnosis -
XX Claim 2; Page 51-52; 187pp; French.
XX The present sequence encodes a protein that is specific for pathogenic
XX strains of Neisseria. The polynucleotides, polypeptides, or their
XX antigenic fragments, are used in vaccines to treat or protect against
XX Neisseria infections, particularly by N. meningitidis. The
XX polynucleotide sequence is also used for recombinant production of
XX the polypeptide and to produce attenuated Neisseria strains that
XX overexpress it, or express it in a non-toxic mutant form.
XX Sequence 858 BP; 297 A; 145 C; 182 G; 234 T; 0 other;
XX
Query Match 98.2%; Score 264.2; DB 21; Length 858;
Best Local Similarity 98.9%; Pred. No. 2.1e-69;
Matches 266; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 CGGAGCATAAATCGTTATTAAAGATAATGTTAGGAACGAGCTTCGATGAAATCAATG 60
Db 158 CGGAGCATAAATCGTTATTCAAGATAATGTTAGGAATGAGCTTCGATGAAATCAATG 217
QY 61 ATTTTATTTCGAAATCGCTCGGAACAGAGGGGAAAGAAACAAAGCTCCCGTGGGAA 120
Db 218 ATTTTATTTCGAAATCGCTCGGAACAGAGGGGAAAGAAACAAAGCTTCCGGTGGGAA 277
QY 121 GAATTCGAACGGGTAAAGAGCGCTTGGTAATTTGGCATTATTTCGGCTTGGCAACAAA 180
Db 278 GAATTCGAACGGGTAAAGAGCGCTTGGTAATTTGGCATTATTTCGGCTTGGCAACAAA 337
QY 181 TTGAATTTCTACTATCCAGGAAACGAAAGGGTTACTTTTACTTTGGATTATGCAGAGA 240
Db 338 TTGAATTTCTACTATCCAGGAAACGAAAGGGTTACTTTTACTTTGGATTATGCAGAGA 397
QY 241 TTCGAAGACGAGGGTATTATCAACCG 269
Db 398 TTCGAAGACGAGGGTATTATCAACCG 426
RESULT 4
ABZ37755
ID ABZ37755 standard; DNA; 855 BP.
XX AC ABZ37755;
XX DT 07-MAR-2003 (first entry)
XX DE N. gonorrhoeae nucleotide sequence SEQ ID 99.
XX KW Antibacterial; infection; vaccine; gene therapy; gene; ds.
XX OS Neisseria gonorrhoeae.
XX PN WO200279243-A2.
XX FD 10-OCT-2002.
XX PF 12-FEB-2002; 2002WO-IB02069.

```

```

XX 12-FEB-2001; 2001GB-0003424.
XX (CHIR-) CHIRON SPA.
XX Fontana MR, Pizza M, Masignani V, Monaci E;
XX WPI; 2003-058415/05.
XX DR P-PSDB; ABP76785.
XX New protein from Neisseria gonorrhoeae, useful for the manufacture of a
XX medicament for treating or preventing N. gonorrhoeae infection -
XX Claim 6; Page 188; 815pp; English.
XX The present invention relates to proteins from Neisseria gonorrhoeae.
XX Also disclosed are the nucleic acid molecules encoding the proteins and
XX antibodies that specifically bind to the proteins. The composition
XX comprising the protein, nucleic acid or antibody is useful for the
XX manufacture of a medicament for treating or preventing N. gonorrhoeae
XX infection, this may be in the form of a vaccine or gene therapy.
XX Sequences given in records ABZ37706-ABZ42016 represent nucleic acid
XX molecules of the invention.
XX Sequence 855 BP; 300 A; 151 C; 178 G; 226 T; 0 other;
XX
Query Match 91.7%; Score 246.6; DB 25; Length 855;
Best Local Similarity 94.8%; Pred. No. 4.1e-64;
Matches 255; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
QY 1 CGGAGCATAAATCGTTATTAAAGATAATGTTAGGAACGAGCTTCGATGAAATCAATG 60
Db 158 CGGAGCATAAATCGTTATTAAAGATAATGTTAGGAATGAGCTTCGATGAAATCAATG 217
QY 61 ATTTTATTTCGAAATCGCTCGGAACAGAGGGGAAAGAAACAAAGCTCCCGTGGGAA 120
Db 218 ATTTTATTTCGAAATCGCTCGGAACAGAGGGGAAAGAAACAAAGCTCCCGTGGGAA 277
QY 121 GAATTCGAACGGGTAAAGAGCGCTTGGTAATTTGGCATTATTTCGGCTTGGCAACAAA 180
Db 278 GAATTCGAACGGGTAAAGAGCGCTTGGTAATTTGGCATTATTTCGGCTTGGCAACAAA 337
QY 181 TTGAATTTCTACTATCCAGGAAACGAAAGGGTTACTTTTACTTTGGATTATGCAGAGA 240
Db 338 TCGAAATCTCTACTATCCAGGAAACGAAAGGGTTACTTTTACTTTGGATTATGCAGAGA 397
QY 241 TTCGAAGACGAGGGTATTATCAACCG 269
Db 398 TTAAGAAAAGTGAAGGTATTATCAACCG 426
RESULT 5
AAA15322
ID AAA15322 standard; DNA; 858 BP.
XX AC AAA15322;
XX DT 04-SEP-2000 (first entry)
XX DE DNA encoding a polypeptide of a Neisseria pathogenic strain.
XX KW Pathogenic strain; Neisseria; vaccine; Neisseria infection; ss.
XX OS Neisseria gonorrhoeae.
XX FH Key Location/Qualifiers
XX CDS 1..858
XX FT /*tag= a
XX PN WO200026375-A2.
XX PD 11-MAY-2000.
XX

```

```

PR 12-FEB-2001; 2001GB-0003424.
XX (CHIR-) CHIRON SPA.
XX Fontana MR, Pizza M, Masignani V, Monaci E;
XX WPI; 2003-058415/05.
DR P-PSDB; ABP80383.
XX New protein from Neisseria gonorrhoeae, useful for the manufacture of a
PT medicament for treating or preventing N. gonorrhoeae infection -
XX Disclosure; Page 718; 815pp; English.
XX The present invention relates to proteins from Neisseria gonorrhoeae.
CC Also disclosed are the nucleic acid molecules encoding the proteins and
CC antibodies that specifically bind to the proteins. The composition
CC comprising the protein, nucleic acid or antibody is useful for the
CC manufacture of a medicament for treating or preventing N. gonorrhoeae
CC infection, this may be in the form of a vaccine or gene therapy.
CC Sequences given in records AB237706-AB242016 represent nucleic acid
CC molecules of the invention.
XX SQ Sequence 891 BP; 316 A; 153 C; 186 G; 236 T; 0 other;
Query Match 91.7%; Score 246.6; DB 25; Length 891;
Best Local Similarity 94.8%; Pred. No. 4.1e-64; Indels 0; Gaps 0;
Matches 255; Conservative 0; Mismatches 14;
QY 1 CGGAGCATAAAATCGTTATTAAAGATAATGTTAGGAACGAGCTTCGATGAATCAATG 60
Db 194 CGGAGCATAAAATCGTTATTAAAGATAATGTCATAGGAATGAGCTTCGATGAATCAATG 253
QY 61 ATTTTATTTCAGAAATCGTTCGGAACAGAGGGAAGAAACAAAGCTCCCGTGGGAA 120
Db 254 ATTTTATTTCAGAAATCGTTCGGAACAGAGGGAAGAAACAAAGCTCCCGTGGGAA 313
QY 121 GAATTCACACGGGTAAAAAGGCTTGGTAAATTTGGCATTTATTCGGGCTTGCCACAAA 180
Db 314 GAATTCACACGGGTAAAAAGGCTTGGTAAATTTGGCATTTATTCAGGCTTGCCACAAA 373
QY 181 TTGAAATTTCTACTATCCAGGGAACGAAAGGTTACTTTTACTTTGGATTATCGAGAGA 240
Db 374 TCGAAATCTCTACTATCCAGGGAACGAAAGGTTACTTTTACTTTGGATTATCGAGAGA 433
QY 241 TTCGAGAGCAAGGTTATTTATCAACCG 269
Db 434 TTAAAAAAGTGAGCGTATTTATCAACCG 462
RESULT 7
AAV03596/c
ID AAV03596 standard; DNA; 270 BP.
XX AC AAV03596;
XX DT 22-OCT-1998 (first entry)
XX DE Neisseria meningitidis DNA sequence E24.
XX KW N. gonorrhoeae; N. lactamica; chromosome Z2491; region 1; region 2;
XX LW region 3; pathogenicity; blood-brain barrier; diagnosis; infection;
XX OS meningitis; ss.
XX OS Neisseria meningitidis.
XX PN WO9802547-A2.
XX PD 22-JAN-1998.
XX PF 11-JUL-1997; 97WO-FR01295.
XX PR 12-JUL-1996; 96FR-0008768.

```

---

```

PF 28-OCT-1999; 99WO-FR02643.
PR 30-OCT-1998; 98FR-0013693.
XX (INRM ) PASTEUR MERIEUX SERUMS & VACCINS SA.
XX (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX Aujaume L, Bouchardon A, Renauld-Mongenie G, Rokbi B, Nassif X;
XX Tinsley C, Ferrin A;
XX WPI; 2000-365622/31.
DR P-PSDB; AAY93292.
XX New polypeptide specific for pathogenic Neisseria useful in therapeutic
PT or preventative vaccines and for diagnosis -
XX Claim 4; Page 134-135; 187pp; French.
XX The present sequence encodes a protein that is specific for pathogenic
CC strains of Neisseria. The polynucleotides, polypeptides, or their
CC antigenic fragments, are used in vaccines to treat or protect against
CC Neisseria infections, particularly by N. meningitidis. The
CC polynucleotide sequence is also used for recombinant production of
CC the polypeptide and to produce attenuated Neisseria strains that
CC overexpress it, or express it in a non-toxic mutant form.
XX SQ Sequence 858 BP; 302 A; 151 C; 178 G; 227 T; 0 other;
Query Match 91.7%; Score 246.6; DB 21; Length 858;
Best Local Similarity 94.8%; Pred. No. 4.1e-64; Indels 0; Gaps 0;
Matches 255; Conservative 0; Mismatches 14;
QY 1 CGGAGCATAAAATCGTTATTAAAGATAATGTTAGGAACGAGCTTCGATGAATCAATG 60
Db 158 CGGAGCATAAAATCGTTATTAAAGATAATGTCATAGGAATGAGCTTCGATGAATCAATG 217
QY 61 ATTTTATTTCAGAAATCGTTCGGAACAGAGGGAAGAAACAAAGCTCCCGTGGGAA 120
Db 218 ATTTTATTTCAGAAATCGTTCGGAACAGAGGGAAGAAACAAAGCTCCCGTGGGAA 277
QY 121 GAATTCACACGGGTAAAAAGGCTTGGTAAATTTGGCATTTATTCGGGCTTGCCACAAA 180
Db 278 GAATTCACACGGGTAAAAAGGCTTGGTAAATTTGGCATTTATTCAGGCTTGCCACAAA 337
QY 181 TTGAAATTTCTACTATCCAGGGAACGAAAGGTTACTTTTACTTTGGATTATCGCAGAGA 240
Db 338 TCGAAATCTCTACTATCCAGGGAACGAAAGGTTACTTTTACTTTGGATTATCGCAGAGA 397
QY 241 TTCGAGAGCAAGGTTATTTATCAACCG 269
Db 398 TTAAAAAAGTGAGCGTATTTATCAACCG 426
RESULT 6
ABZ41353
ID ABZ41353 standard; DNA; 891 BP.
XX AC ABZ41353;
XX DT 07-MAR-2003 (first entry)
XX DE N. gonorrhoeae nucleotide sequence SEQ ID 7295.
XX KW Antibacterial; infection; vaccine; gene therapy; gene; ds.
XX OS Neisseria gonorrhoeae.
XX PN WO200279243-A2.
XX PD 10-OCT-2002.
XX PF 12-FEB-2002; 2002WO-IB02069.
XX PR

```



us-09-928-457-77.rng

Sun Nov 16 12:56:59 2003

PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226888.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 12-SEP-2000; 2000US-0232081.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.

PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 20-OCT-2000; 2000US-0242221.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250391.  
PR 01-DEC-2000; 2000US-0251160.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0255719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-02559678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-541565/60.

Nucleic acids encoding 3224 human nervous system antigen polypeptides,  
useful for preventing, diagnosing and/or treating nervous system  
cancers and metastases -

Disclosure; SEQ ID NO 11501; 1701bp + Sequence Listing; English.

XX



PR	18-APR-2000;	2000US-01198112;
PR	19-MAY-2000;	2000US-02059515;
PR	07-JUN-2000;	2000US-02094617;
PR	28-JUN-2000;	2000US-02148886;
PR	30-JUN-2000;	2000US-02151535;
PR	07-JUL-2000;	2000US-02166477;
PR	07-JUL-2000;	2000US-02168800;
PR	11-JUL-2000;	2000US-02174480;
PR	11-JUL-2000;	2000US-02174936;
PR	14-JUL-2000;	2000US-02184290;
PR	25-JUL-2000;	2000US-02209633;
PR	25-JUL-2000;	2000US-02209664;
PR	14-AUG-2000;	2000US-02245158;
PR	14-AUG-2000;	2000US-02245193;
PR	14-AUG-2000;	2000US-02252313;
PR	14-AUG-2000;	2000US-02252314;
PR	14-AUG-2000;	2000US-02252365;
PR	14-AUG-2000;	2000US-02252627;
PR	14-AUG-2000;	2000US-02252770;
PR	14-AUG-2000;	2000US-02254477;
PR	14-AUG-2000;	2000US-02255757;
PR	14-AUG-2000;	2000US-02257558;
PR	14-AUG-2000;	2000US-02257559;
PR	18-AUG-2000;	2000US-02262379;
PR	22-AUG-2000;	2000US-02266811;
PR	22-AUG-2000;	2000US-02268668;
PR	22-AUG-2000;	2000US-02271882;
PR	23-AUG-2000;	2000US-02277009;
PR	30-AUG-2000;	2000US-02289224;
PR	01-SEP-2000;	2000US-02292897;
PR	01-SEP-2000;	2000US-02293443;
PR	01-SEP-2000;	2000US-02293444;
PR	01-SEP-2000;	2000US-02293445;
PR	05-SEP-2000;	2000US-02295059;
PR	05-SEP-2000;	2000US-02295093;
PR	06-SEP-2000;	2000US-02304337;
PR	06-SEP-2000;	2000US-02304338;
PR	08-SEP-2000;	2000US-02320281;
PR	12-SEP-2000;	2000US-02319668;
PR	14-SEP-2000;	2000US-02329377;
PR	14-SEP-2000;	2000US-02329387;
PR	14-SEP-2000;	2000US-02329399;
PR	14-SEP-2000;	2000US-02324000;
PR	14-SEP-2000;	2000US-02324011;
PR	14-SEP-2000;	2000US-02330633;
PR	14-SEP-2000;	2000US-02330664;
PR	14-SEP-2000;	2000US-02330665;
PR	21-SEP-2000;	2000US-02342233;
PR	21-SEP-2000;	2000US-02342774;
PR	25-SEP-2000;	2000US-02349597;
PR	25-SEP-2000;	2000US-02349598;
PR	25-SEP-2000;	2000US-02354884;
PR	29-SEP-2000;	2000US-02363707;
PR	02-OCT-2000;	2000US-02368082;
PR	02-OCT-2000;	2000US-02370737;
PR	02-OCT-2000;	2000US-02370738;
PR	02-OCT-2000;	2000US-02370739;
PR	13-OCT-2000;	2000US-02370740;
PR	13-OCT-2000;	2000US-02399935;
PR	13-OCT-2000;	2000US-02399937;

PR	12-SEP-2000	2000US-0231968
PR	14-SEP-2000	2000US-0232397
PR	14-SEP-2000	2000US-0232398
PR	14-SEP-2000	2000US-0232399
PR	14-SEP-2000	2000US-0232400
PR	14-SEP-2000	2000US-0232401
PR	14-SEP-2000	2000US-0232406
PR	14-SEP-2000	2000US-0233064
PR	14-SEP-2000	2000US-0233065
PR	21-SEP-2000	2000US-0234223
PR	21-SEP-2000	2000US-0234274
PR	25-SEP-2000	2000US-0234997
PR	25-SEP-2000	2000US-0234998
PR	26-SEP-2000	2000US-0235484
PR	27-SEP-2000	2000US-0235834
PR	27-SEP-2000	2000US-0235836
PR	29-SEP-2000	2000US-0236327
PR	29-SEP-2000	2000US-0236367
PR	29-SEP-2000	2000US-0236368
PR	29-SEP-2000	2000US-0236369
PR	29-SEP-2000	2000US-0236370
PR	02-OCT-2000	2000US-0236802
PR	02-OCT-2000	2000US-0237037
PR	02-OCT-2000	2000US-0237038
PR	02-OCT-2000	2000US-0237039
PR	02-OCT-2000	2000US-0237040
PR	13-OCT-2000	2000US-0239935
PR	13-OCT-2000	2000US-0239937

20-OCT-2000; 2000US-0240960.  
20-OCT-2000; 2000US-0241221.  
20-OCT-2000; 2000US-0241785.  
20-OCT-2000; 2000US-0241786.  
20-OCT-2000; 2000US-0241787.  
20-OCT-2000; 2000US-0241808.  
20-OCT-2000; 2000US-0241809.  
20-OCT-2000; 2000US-0241826.  
20-OCT-2000; 2000US-0241827.  
01-NOV-2000; 2000US-0246474.  
08-NOV-2000; 2000US-0246475.  
08-NOV-2000; 2000US-0246476.  
08-NOV-2000; 2000US-0246477.  
08-NOV-2000; 2000US-0246478.  
08-NOV-2000; 2000US-0246523.  
08-NOV-2000; 2000US-0246524.  
08-NOV-2000; 2000US-0246525.  
08-NOV-2000; 2000US-0246526.  
08-NOV-2000; 2000US-0246527.  
08-NOV-2000; 2000US-0246528.  
08-NOV-2000; 2000US-0246532.  
08-NOV-2000; 2000US-0246609.  
08-NOV-2000; 2000US-0246610.  
08-NOV-2000; 2000US-0246611.  
08-NOV-2000; 2000US-0246613.  
17-NOV-2000; 2000US-0249207.  
17-NOV-2000; 2000US-0249208.  
17-NOV-2000; 2000US-0249209.  
17-NOV-2000; 2000US-0249210.  
17-NOV-2000; 2000US-0249211.  
17-NOV-2000; 2000US-0249212.  
17-NOV-2000; 2000US-0249213.  
17-NOV-2000; 2000US-0249214.  
17-NOV-2000; 2000US-0249215.  
17-NOV-2000; 2000US-0249216.  
17-NOV-2000; 2000US-0249217.  
17-NOV-2000; 2000US-0249218.  
17-NOV-2000; 2000US-0249244.  
17-NOV-2000; 2000US-0249245.  
17-NOV-2000; 2000US-0249264.  
17-NOV-2000; 2000US-0249265.  
17-NOV-2000; 2000US-0249297.  
17-NOV-2000; 2000US-0249299.  
17-NOV-2000; 2000US-0249300.  
01-DEC-2000; 2000US-0250160.  
01-DEC-2000; 2000US-0250391.  
05-DEC-2000; 2000US-0251030.  
05-DEC-2000; 2000US-0251989.  
05-DEC-2000; 2000US-0251989.  
05-DEC-2000; 2000US-0251990.  
11-DEC-2000; 2000US-0254097.  
05-JAN-2001; 2001US-0259676.  
(HUMA-) HUMAN GENOME SCI INC.  
Rosen CA, Bazash SC, Ruben SM;  
WPI; 2001-483426/52.  
Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
useful for preventing, diagnosing and/or treating cancers and  
metastasis -  
Disclosure; SEQ ID NO 36112; 3071pp + Sequence Listing; English.  
AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)  
amino acid sequences given in AAK54951 to AAK64702. (I) have cytostatic  
activity, and can be used in gene therapy and vaccine production. (I)

CC proteins and polynucleotides may be used in the prevention, diagnosis and  
CC treatment of diseases associated with inappropriate (I) expression. For  
CC example, they may be used to treat disorders associated with decreased  
CC expression by rectifying mutations or deletions in a patient's genome  
CC that affect the activity of (I) by expressing inactive proteins or to  
CC supplement the patient's own production of (I). Additionally, (I)  
CC polynucleotides may be used to produce the secreted (I), by inserting the  
CC the nucleic acids into a host cell and culturing the cell to express the  
CC protein. (I) proteins and polynucleotides may be used to prevent,  
CC diagnose and treat immune/hematopoietic-related diseases, especially  
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703  
CC to AAK87694 represent human immune/hematopoietic antigen genomic  
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169  
CC represent sequences used in the exemplification of the present invention.  
XX  
SQ Sequence 13443 BP; 4183 A; 2486 C; 2503 G; 4271 T; 0 other;  
Query Match 15.2%; Score 41; DB 22; Length 13443;  
Best Local Similarity 48.1%; Fred. No. 0.045;  
Matches 116; Conservative 0; Mismatches 125; Indels 0; Gaps 0;  
QY 2 GGAGCATATAAATCGTTATTAAAGATAATGTTATAGGACGAGCTTCGATGAAATCAATGA 61  
DB 6608 GTAGCATTTGATCATTTGACCTTCAATGATACGATTGGCTTACTACAGAGGAAAG 6549  
QY 62 TTTTATTGAGATCGGTGGACAGAGGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 121  
DB 6548 CAATTGGTGTCTCTATGCAAAATTTATTAAGTAACCAAGGAGGAGGAGGAGGAGGAG 6489  
QY 122 AATTCACAGGGGTAAAAAGGCTTGTAAATTTGGCAATTTTCGGCTTGCACAAAAAT 181  
DB 6488 GAGTCTACAGGAGATTTCTTCTAGCTCAGTGTGACAGTATGCTACAGATCAGG 6429  
QY 182 TCAAAATTTCTACTATCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 241  
DB 6428 AGGGATTTAGTGTTCCTCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 6369  
QY 242 T 242  
DB 6368 T 6368  
RESULT 11  
ABA19169/c  
ID ABA19169 standard; DNA; 13451 BP.  
XX ABA19169;  
XX  
DT 23-JAN-2002 (first entry)  
XX  
DE Human nervous system related polynucleotide SEQ ID NO 11500.  
XX  
XX Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;  
XX immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnary;  
XX antiparkinsonian; antischizoid; antianemic; antiarthritic; cancer;  
XX antitubercular; hepatotropic; cerebroprotective; antiinflammatory;  
XX antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;  
XX antiparasitic; cardiac; immune disorder; cardiovascular disorder;  
XX neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200159063-A2.  
XX  
PD 16-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01334.  
XX  
XX 31-JAN-2000; 2000US-0179065.  
XX 04-FEB-2000; 2000US-0180628.  
XX 24-FEB-2000; 2000US-0184664.  
XX 02-MAR-2000; 2000US-0186350.  
XX 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190075.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226688.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235835.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 20-OCT-2000; 2000US-0242221.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249266.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 03-JAN-2001; 2001US-0259678.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-541565/60.  
XX  
XX Nucleic acids encoding 324 human nervous system antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating nervous system  
PT cancers and metastases -  
XX  
XX Disclosure; SEQ ID NO 11500; 1701pp + Sequence Listing; English.  
PS  
XX  
XX The invention relates to novel genes (ABA11004-ABA21534) and proteins  
CC (ABBI4678-ABBI8001) useful for preventing, treating or ameliorating  
CC

CC medical conditions e.g. by protein or gene therapy. The genes are  
 CC isolated from a range of human tissues disclosed in the specification.  
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful  
 CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone  
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;  
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
 CC and parasitic infections.  
 CC Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 13451 BP; 4193 A; 2486 C; 2502 G; 4270 T; 0 other;

Query Match 14.6%; Score 39.4; DB 22; Length 13451;  
 Best Local Similarity 47.7%; Pred. No. 0.14;  
 Matches 115; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

Qy 2 GGACATAAATCGTTTATTAAAGATATGGTATAGGAACGAGCTTCGATGAATCAATGA 61  
 Db 6612 GTAGCATTTGATCAATTCACCTTCAATGATGATGGATGGCTTACTACAGAGGAAAG 6553

Qy 62 TTTTATTGAGATCGTCCGGAACAGAGGGAAGAAACACGCTCCCGTGGGAAG 121  
 Db 6552 CAATGGTGCTTATGCAAAATTTATATAGTAACACAGAAATGGCTACAGTGGGATGA 6493

Qy 122 AATTCACAGGGTAAAGAGCCCTTGTAATATGSCATTTATCGGCTTGGCAACAAAAT 181  
 Db 6492 GAGTCCTACAGGAAGATTTCTTCTAGCTCAGTTGGTACAGTATGTGACTCAGAAATCAGG 6433

Qy 182 TGAATTTCTACTCCAGGGAACAGAGGTTTACTTTTACTTTGGATTATGCAGAT 241  
 Db 6432 AGGATTTAGTGTTCACAGAAACACAGATTTTGTCATGCTTAGAATATAGATATAT 6373

Qy 242 T 242  
 Db 6372 T 6372

RESULT 12  
 AAK81299/c  
 ID AAK81299 standard; DNA; 13451 BP.  
 XX  
 AC AAK81299;  
 XX  
 DT 07-NOV-2001 (first entry)  
 XX  
 DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:36111.  
 XX  
 KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
 KW cytostatic; gene therapy; vaccine; metastasis; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 FN WO200157182-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PP 17-JAN-2001; 2001WO-US01354.  
 XX  
 PP 31-JAN-2000; 2000US-0179065.  
 PR 04-FEB-2000; 2000US-0180628.  
 PR 24-FEB-2000; 2000US-0184664.  
 PR 02-MAR-2000; 2000US-0186350.  
 PR 16-MAR-2000; 2000US-0189874.  
 PR 17-MAR-2000; 2000US-0190076.  
 PR 18-APR-2000; 2000US-0198123.  
 PR 19-MAY-2000; 2000US-0205515.  
 PR 07-JUN-2000; 2000US-0209467.  
 PR 28-JUN-2000; 2000US-0214886.  
 PR 30-JUN-2000; 2000US-0215135.  
 PR 07-JUL-2000; 2000US-0216647.  
 PR 07-JUL-2000; 2000US-0216880.  
 PR 11-JUL-2000; 2000US-0217487.  
 PR 11-JUL-2000; 2000US-0217496.  
 PR 14-JUL-2000; 2000US-0218290.  
 PR 26-JUL-2000; 2000US-0220963.  
 PR 26-JUL-2000; 2000US-0220964.  
 PR 14-AUG-2000; 2000US-0224518.  
 PR 14-AUG-2000; 2000US-0224519.  
 PR 14-AUG-2000; 2000US-0225213.  
 PR 14-AUG-2000; 2000US-0225214.  
 PR 14-AUG-2000; 2000US-0225266.  
 PR 14-AUG-2000; 2000US-0225267.  
 PR 14-AUG-2000; 2000US-0225268.  
 PR 14-AUG-2000; 2000US-0225270.  
 PR 14-AUG-2000; 2000US-0225447.  
 PR 14-AUG-2000; 2000US-0225757.  
 PR 14-AUG-2000; 2000US-0225758.  
 PR 14-AUG-2000; 2000US-0225759.  
 PR 18-AUG-2000; 2000US-0226279.  
 PR 22-AUG-2000; 2000US-0226681.  
 PR 22-AUG-2000; 2000US-0226688.  
 PR 22-AUG-2000; 2000US-0227182.  
 PR 23-AUG-2000; 2000US-0227009.  
 PR 30-AUG-2000; 2000US-0228924.  
 PR 01-SEP-2000; 2000US-0229287.  
 PR 01-SEP-2000; 2000US-0229343.  
 PR 01-SEP-2000; 2000US-0229344.  
 PR 01-SEP-2000; 2000US-0229345.  
 PR 05-SEP-2000; 2000US-0229509.  
 PR 05-SEP-2000; 2000US-0229513.  
 PR 06-SEP-2000; 2000US-0230437.  
 PR 06-SEP-2000; 2000US-0230438.  
 PR 08-SEP-2000; 2000US-0231242.  
 PR 08-SEP-2000; 2000US-0231243.  
 PR 08-SEP-2000; 2000US-0231244.  
 PR 08-SEP-2000; 2000US-0231413.  
 PR 08-SEP-2000; 2000US-0231414.  
 PR 08-SEP-2000; 2000US-0232080.  
 PR 08-SEP-2000; 2000US-0232081.  
 PR 12-SEP-2000; 2000US-0231968.  
 PR 14-SEP-2000; 2000US-0232397.  
 PR 14-SEP-2000; 2000US-0232398.  
 PR 14-SEP-2000; 2000US-0232399.  
 PR 14-SEP-2000; 2000US-0232400.  
 PR 14-SEP-2000; 2000US-0232401.  
 PR 14-SEP-2000; 2000US-0233063.  
 PR 14-SEP-2000; 2000US-0233064.  
 PR 14-SEP-2000; 2000US-0233065.  
 PR 21-SEP-2000; 2000US-0234223.  
 PR 21-SEP-2000; 2000US-0234274.  
 PR 25-SEP-2000; 2000US-0234997.  
 PR 25-SEP-2000; 2000US-0234998.  
 PR 26-SEP-2000; 2000US-0235484.  
 PR 27-SEP-2000; 2000US-0235834.  
 PR 27-SEP-2000; 2000US-0235835.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236367.  
 PR 29-SEP-2000; 2000US-0236368.  
 PR 29-SEP-2000; 2000US-0236369.  
 PR 29-SEP-2000; 2000US-0236370.  
 PR 02-OCT-2000; 2000US-0236802.  
 PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237038.  
 PR 02-OCT-2000; 2000US-0237039.  
 PR 13-OCT-2000; 2000US-0239935.  
 PR 13-OCT-2000; 2000US-0239937.  
 PR 20-OCT-2000; 2000US-0240960.  
 PR 20-OCT-2000; 2000US-0241221.

20-OCT-2000; 2000US-0241785.  
20-OCT-2000; 2000US-0241786.  
20-OCT-2000; 2000US-0241787.  
20-OCT-2000; 2000US-0241808.  
20-OCT-2000; 2000US-0241809.  
20-OCT-2000; 2000US-0241826.  
01-NOV-2000; 2000US-0244617.  
08-NOV-2000; 2000US-0246474.  
08-NOV-2000; 2000US-0246475.  
08-NOV-2000; 2000US-0246476.  
08-NOV-2000; 2000US-0246477.  
08-NOV-2000; 2000US-0246478.  
08-NOV-2000; 2000US-0246523.  
08-NOV-2000; 2000US-0246524.  
08-NOV-2000; 2000US-0246525.  
08-NOV-2000; 2000US-0246526.  
08-NOV-2000; 2000US-0246527.  
08-NOV-2000; 2000US-0246528.  
08-NOV-2000; 2000US-0246532.  
08-NOV-2000; 2000US-0246609.  
08-NOV-2000; 2000US-0246610.  
08-NOV-2000; 2000US-0246611.  
08-NOV-2000; 2000US-0246613.  
17-NOV-2000; 2000US-0249207.  
17-NOV-2000; 2000US-0249208.  
17-NOV-2000; 2000US-0249209.  
17-NOV-2000; 2000US-0249210.  
17-NOV-2000; 2000US-0249211.  
17-NOV-2000; 2000US-0249212.  
17-NOV-2000; 2000US-0249213.  
17-NOV-2000; 2000US-0249214.  
17-NOV-2000; 2000US-0249215.  
17-NOV-2000; 2000US-0249216.  
17-NOV-2000; 2000US-0249217.  
17-NOV-2000; 2000US-0249218.  
17-NOV-2000; 2000US-0249244.  
17-NOV-2000; 2000US-0249245.  
17-NOV-2000; 2000US-0249264.  
17-NOV-2000; 2000US-0249265.  
17-NOV-2000; 2000US-0249297.  
17-NOV-2000; 2000US-0249299.  
17-NOV-2000; 2000US-0249300.  
01-DEC-2000; 2000US-0250160.  
01-DEC-2000; 2000US-0250391.  
05-DEC-2000; 2000US-0251030.  
05-DEC-2000; 2000US-0251988.  
05-DEC-2000; 2000US-0256719.  
06-DEC-2000; 2000US-0251479.  
08-DEC-2000; 2000US-0251856.  
08-DEC-2000; 2000US-0251868.  
08-DEC-2000; 2000US-0251869.  
08-DEC-2000; 2000US-0251989.  
08-DEC-2000; 2000US-0251990.  
11-DEC-2000; 2000US-0254097.  
05-JAN-2001; 2001US-02559678.  
(HUMA-) HUMAN GENOME SCI INC.  
Rosen CA, Barash SC, Ruben SM;  
WPI; 2001-483426/52.  
Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
useful for preventing, diagnosing and/or treating cancers and  
metastasis -  
Disclosure; SEQ ID NO 36111; 3071pp + Sequence Listing; English.  
AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)  
amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic  
activity, and can be used in gene therapy and vaccine production. (I)  
proteins and polynucleotides may be used in the prevention, diagnosis and  
treatment of diseases associated with inappropriate (I) expression. For

CC example, they may be used to treat disorders associated with decreased  
CC expression by rectifying mutations or deletions in a patient's genome  
CC that affect the activity of (I) by expressing inactive proteins or to  
CC supplement the patient's own production of (I). Additionally, (I)  
CC polynucleotides may be used to produce the secreted (I), by inserting  
CC the nucleic acids into a host cell and culturing the cell to express the  
CC protein. (I) proteins and polynucleotides may be used to prevent, the  
CC diagnose and treat immune/hematopoietic-related diseases, especially  
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703  
CC to AAK87694 represent human immune/hematopoietic antigen genomic  
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169  
CC represent sequences used in the exemplification of the present invention.  
XX  
SQ Sequence 13451 BP; 4193 A; 2486 C; 2502 G; 4270 T; 0 other;  
Query Match 14.6%; Score 39.4; DB 22; Length 13451;  
Best Local Similarity 47.7%; Pred. No. 0.14;  
Matches 115; Conservative 0; Mismatches 126; Indels 0; Gaps 0;  
QY 2 GGAGCATAAATCGTTATTAAAGATAATGGTATAGGACGAGCTTCGATGAATCAATCA 61  
Db 6612 GTAGCATTTGATCAITGACTTCAATGATGACGTTGGTGGCTTACTCAGAGGAAAG 6553  
QY 62 TTTTATTGTAGATCGGTGCGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 121  
Db 6552 CAATTGGTCTCTATGCAAAATTTATTAAGTAACACGAAATGGCTACAGTGGATGA 6493  
QY 122 AATTCACAGGGTAAAGAGGCTTGGTAAATGTCATATTCGGGCTTGGCAACAAAT 181  
Db 6492 GAGTCTACAGGAGGATTCCTCCTAGCTCAGTGTGTCAGTATGTGACTCAGAAATCAGG 6433  
QY 182 TGAATTTCTACTATCCAGGGAACGAAAGGCTTACTTTTACTTTGGATTATGCAGAGAT 241  
Db 6432 AGGATTTAGTGTTCCTCCAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 241  
QY 242 T 242  
Db 6372 T 6372  
RESULT 13  
ABU13122  
ID ABL13122 standard; cDNA; 4951 BP.  
XX  
AC ABL13122;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 33848.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ss.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
DR WPI; 2001-656860/75.  
DR P-PSDB; ABB69019.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions -  
PS Claim 1; SEQ ID NO 33848; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from *Drosophila*. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 4951 BP; 1376 A; 1121 C; 1065 G; 1389 T; 0 other;  
Query Match 13.7%; Score 36.8; DB 23; Length 4951;  
Best Local Similarity 61.5%; Pred. No. 0.62;  
Matches 59; Conservative 0; Mismatches 37; Indels 0; Gaps 0;  
QY 1 CGGAGCATAAATCGTTTATTAAAGATAATGTTATAGGACGAGCTTCGATGAATCAATG 60  
DB 3962 CCGAGCATCAATCGTTTATTAGTTATATGTTAAAGAGATAATTTGTTTAAATATATG 4021  
QY 61 ATTTTATTTCAGAAATCGTTCGGAACAGAGGGAAG 96  
DB 4022 ATATATTTTGTAGTATACGGTTACAGAGTGCATG 4057  
RESULT 14  
ABL15458  
ID ABL15458 standard; cDNA; 5395 BP.  
XX  
AC ABL15458;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE *Drosophila melanogaster* expressed polynucleotide SEQ ID NO 40856.  
XX  
KW *Drosophila*; developmental biology; cell signalling; insecticide;  
XX pharmaceutical; gene; ss.  
XX  
OS *Drosophila melanogaster*.  
XX WO200171042-A2.  
PN  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
DR WPI; 2001-656860/75.  
DR P-PSDB; ABB71355.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from *Drosophila* and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
FS Claim 1; SEQ ID NO 40856; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from *Drosophila*. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 5395 BP; 1528 A; 1109 C; 1141 G; 1617 T; 0 other;  
Query Match 13.7%; Score 36.8; DB 23; Length 5395;  
Best Local Similarity 61.5%; Pred. No. 0.63;  
Matches 59; Conservative 0; Mismatches 37; Indels 0; Gaps 0;  
QY 1 CGGAGCATAAATCGTTTATTAAAGATAATGTTATAGGACGAGCTTCGATGAATCAATG 60  
DB 441 CCGAGCATCAATCGTTTATTAGTTATATGTTAAAGAGATAATTTGTTTAAATATATG 500  
QY 61 ATTTTATTTCAGAAATCGTTCGGAACAGAGGGAAG 96  
DB 501 ATATATTTTGTAGTATACGGTTACAGAGTGCATG 536  
RESULT 15  
ABL216977/c  
ID ABL216977 standard; DNA; 2000 BP.  
XX  
AC ABL216977;  
XX  
DT 21-JAN-2003 (first entry)  
XX  
DE *Arabidopsis thaliana* stress regulated gene SEQ ID NO 4782.  
XX  
KW *Arabidopsis thaliana*; plant; gene; stress; transgenic; ds.  
XX  
OS *Arabidopsis thaliana*.  
XX WO200216655-A2.  
PN  
PD 28-FEB-2002.  
XX  
PF 24-AUG-2001; 2001WO-US26685.  
XX  
PR 24-AUG-2000; 2000US-227866P.  
PR 26-JAN-2001; 2001US-264647P.  
PR 22-JUN-2001; 2001US-300111P.  
XX  
PA (SCRI ) SCRIPPS RES INST.  
PA (SYGN ) SYNGENTA PARTICIPATIONS AG.  
XX  
PI Harper JF, Krepis J, Wang X, Zhu T;  
XX  
DR WPI; 2002-304127/34.  
XX  
PT Identifying a stress condition to which a plant cell has been exposed  
PT and producing plants with increased tolerance to these abiotic stresses  
PT -  
XX  
PS Claim 144; SEQ ID NO 4782; 577pp + Sequence Listing; English.  
XX  
CC The invention relates to identifying a stress condition to which a plant  
CC cell has been exposed, comprising:  
CC (a) contacting nucleic acid representative of expressed polynucleotides  
CC in the plant cell with an array or probes representative of the plant  
CC cell genome; and  
CC (b) detecting a profile of expressed polynucleotides in the plant cell  
CC characteristic of a stress response. The method is useful in the  
CC production of transgenic plants, cells and seeds and in producing plants  
CC with increased tolerance to abiotic stress. The present sequence is that  
CC of an *Arabidopsis thaliana* stress regulated gene (ABL21696-ABL217574) used  
CC in methods of the invention.  
CC Note: The sequence data for this patent is not represented in the printed  
CC specification but is based on sequence information supplied to Derwent by  
CC the European Patent Office.

Search completed: November 15, 2003, 00:35:27  
Job time : 176.613 secs







```

FEATURES
  source
    pBelobAC11.
      Location/Qualifiers
        1..1101
          /organism="Drosophila melanogaster"
          /mol_type="genomic DNA"
          /db_xref="taxon:7227"
          /clone="BACN16L21"
          /clone_lib="DrosBAC"
          /plasmid="pBelobAC11"
          /note="end : T7"
BASE COUNT      402 a 153 c 67 g 213 t 266 others
ORIGIN
Query Match      15.1%; Score 40.6; DB 29; Length 1101;
Best Local Similarity 32.5%; Pred. No. 2.3;
Matches 79; Conservative 64; Mismatches 99; Indels 1; Gaps 1;
QY 16 TTATTAAGATAAATGGTATAGGACGAGCTTCGATGAATCAATGATTTTATTTCGAGAA 75
Db 823 TTTTSTAWATAAAMCASAAMASVASSTTCSAWTAALAAWKTWTTAAATATATGA 882
QY 76 TCGTCCGAAACAGAGGAGAAACAAAGCCTCCCGTCGCGGAAGAAATCCAAAGGTA 135
Db 883 TAGCGGGGAAATAACGAKASSDRAACGSSSGGAATKASAAVATTATTSACWSTAAA 942
QY 136 AAAAAGCCCTGGTAATTCGCATATTCGGGCTTGGCAAC-AAATTTGAAATTTCTACT 194
Db 943 AADADSKSSGTFKWSGTGGKRTTATTADACBMAAACCCCBTAACBKAACACCMWMA 1002
QY 195 ATCCAGGAAACGAAAGGTTACTTTTACCTTTGATTATGCGAGATTCGAAAGAACGAG 254
Db 1003 TTTWAAAAAWAMACVTTAASTWAAVTTTBTATSSGVBSTSSMSDTSSTSSS 1062
QY 255 GGT 257
Db 1063 SST 1065

RESULT 2
AL546670      1201 bp mRNA linear EST 31-MAY-2003
LOCUS      AL546670 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cdNA
DEFINITION      clone CS0D1029YD16 5-PRIME, mRNA sequence.
ACCESSION      AL546670
VERSION      AL546670.2 GI:31268503
KEYWORDS      EST.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
REFERENCE      1 (bases 1 to 1201)
AUTHORS      Li W.B., Gruber C., Jesses, J. and Polayes, D.
TITLE      Full-length cDNA libraries and normalization
COMMENT      Unpublished
      On Feb 15, 2001 this sequence version replaced gi:12880010.
      Contact: Genoscope
      Genoscope - Centre National de Sequencage
      BP 191 91006 EVRY cedex - France
      Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
      Library was constructed by Life Technologies, a division of
      Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
      http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
      Faraday Avenue Genoscope sequence ID : CS0D1029YD08QF1.
      Location/Qualifiers
        1..1201
          /organism="Homo sapiens"
          /mol_type="mRNA"
          /db_xref="taxon:9606"
          /clone="CS0D1029YD16"
          /tissue_type="PLACENTA COT 25-NORMALIZED"
          /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
          /note="1st strand cDNA was primed with a NotI-oligo(dT)
          primer. Five prime end enriched, double-strand cDNA was
          digested with Not I and cloned into the Not I and EcoR V
          sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT      302 a 255 c 226 g 296 t 122 others
ORIGIN
Query Match      14.6%; Score 39.2; DB 9; Length 1201;
Best Local Similarity 36.4%; Pred. No. 5.7;
Matches 68; Conservative 37; Mismatches 82; Indels 0; Gaps 0;
QY 63 TTTTATTTCGATCGTCGGAACAGAGGAGAAACAAAGCCTCCCGTCGCGGAAGA 122
Db 915 WTTTTTTTTTAAAGGKYCYVTWAAWKTWTTWAAAAAAGVMDGSSGKTTTAA 974
QY 123 ATCCACGGGTAAAGAGCCTTGGTAAATTCGCATATTCGGGCTTGGCAACAAAT 182
Db 975 AAKGARKARGAAAVANWAKTTTGGKTTTTTTAAARWCTHTTTTGGGGRAAAATG 1034
QY 183 GAAATTTCTACTATCCAGGGAACGAAAGGTTTACTTTTACTTTTGGATTATCGAGATT 242
Db 1035 RAAAAAMCWAAGTTTHMSBYAAGGAWAWTTTITDKTTTTTAAAAAASAAHBS 1094
QY 243 CGAAGAA 249
Db 1095 GSKGAA 1101

RESULT 3
T09518/c
LOCUS      T09518 289 bp DNA linear GSS 05-DEC-2000
DEFINITION      clone 0113m, G. Roman Reddy Plasmodium falciparum genomic
ACCESSION      T09518
VERSION      T09518.1 GI:319350
KEYWORDS      GSS.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
ORGANISM      Plasmodium falciparum
REFERENCE      1 (bases 1 to 289)
AUTHORS      Reddy, G.R., Chakrabarti, D., Schuster, S.M., Ferl, R.J., Almira, E.C.
TITLE      Gene sequence tags from Plasmodium falciparum genomic DNA fragments
      prepared by the genome activity of mung bean nuclease
JOURNAL      Proc. Natl. Acad. Sci. U.S.A. 90, 9867-9871 (1993)
MEDLINE      94052193
PUBMED      8234327
COMMENT      Other GSSs: 0113m3
      Contact: Dame JB
      Dept. of Pathobiology, College of Veterinary Medicine
      University of Florida
      2015 SW 23rd Avenue, Bldg 1017, Gainesville, FL 32611, USA
      Tel: 352 392 4700
      Fax: 352 392 9704
      Email: damej@mail.vetmed.ufl.edu
      Seq primer: T7
      Class: Shotgun.
      Location/Qualifiers
        1..289
          /organism="Plasmodium falciparum"
          /mol_type="genomic DNA"
          /db_xref="taxon:5833"
          /clone="0113m"
          /lab_host="E. coli XL1-Blue"
          /clone_lib="gmbpHB3.1, G. Roman Reddy"
          /note="Vector: pBluescript SK(+); Genomic DNA, from
          asynchronous blood stage parasites of the cloned Honduran
          HB3 isolate cultured in vitro, was digested with mung bean
          nuclease in the presence of 30% formamide at 50°C (Vernick
          K.D., Imberski, R.B., and McCutchan, T.F. 1988. Nucleic
          Acids Research 16:683-689). The ends of the fragments
          were polished using T4 DNA polymerase, and the fragments
          were ligated to EcoR V-cleaved and dephosphorylated
          pBluescript SK(+). Recombinant plasmids transformed E.
          coli XL1-Blue."

```

```

BASE COUNT      117 a      29 c      31 g      100 t      12 others
ORIGIN
Query Match      14.4%; Score 38.8; DB 29; Length 289;
Best Local Similarity 48.4%; Pred. No. 5.8;
Matches 91; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

QY 4 AGCATAAATCGTTATTAAAGATAATGTTATAGAACGAGCTTCGATGAATCAATGATT 63
    |||||
DB 278 AGCANTGANNATNATATATAAGAAATANTATTGGTGAAATATTATTTTGATGATTATC 219
    |||||
--QY 64 TTTATTTGAGATCGTTCGGAACAGAGGAGAGAAAACAAGCTCCCGTGGCGAAGAA 123
    |||||
DB 218 NTTTTTCAAATCATTTATGTAATATGTAATGTTATTTAACTTACCTTCAGATGCA 159
    |||||
QY 124 TTCCACGGGTAAAAAAGCGCTTGTAATTTGGCATTTTCGGCTTGGCAACAAAATTG 183
    |||||
DB 158 CTAANAATTTAAAAAGTCTTCATATTAACNACATATATTTCGTCCTCAATCCCTTT 99
    |||||
QY 184 AATTTTCT 191
    |||||
DB 98 AATTTT 91
    |||||

RESULT 4
BX437575/c
LOCUS      BX437575 Homo sapiens THYMUS Homo sapiens cDNA clone CSOCAP007YF09
DEFINITION 3-PRIME, mRNA sequence.
ACCESSION  BX437575
VERSION     BX437575.1 GI:30771584
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens

REFERENCE   1 (bases 1 to 797)
AUTHORS    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE       Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL     Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
COMMENT     Full-length cDNA libraries and normalization
            Unpublished
            Contact: Genoscope
            Genoscope - Centre National de Séquençage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 10594.f For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CSOCAP007CC05NP1&cluster=10594.f. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CSOCAP007CC05NP1.
            Location/Qualifiers
                1..797
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="CSOCAP007YF09"
                /tissue type="THYMUS"
                /clone_lib="Homo sapiens THYMUS"
                /note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
                with a NotI-oligo(dT) primer. Five prime end enriched,
                double-strand cDNA was digested with Not I and cloned into
                the Not I and EcoRV sites of the pCMVSPORT 6 vector.
                Library was not normalized."
                BASE COUNT      101 a      142 c      102 g      288 t      164 others
ORIGIN

Query Match      14.3%; Score 38.6; DB 13; Length 797;
Best Local Similarity 26.7%; Pred. No. 7.7;
Matches 65; Conservative 75; Mismatches 103; Indels 0; Gaps 0;

QY 7 ATAAATCGTTATTAAAGATAATGTTATAGAACGAGCTTCGATGAATCAATGATTTT 66
    |||||
DB 153 AAAAAAARGTAAAMMAAAAAAARAAAAAARAAAAAARAAAAAARAAAAAARAAAAA 212
    |||||
QY 7 ATAAATCGTTATCGTCGGAACAGAGGAAGAAAAACAAGCCTCCCGTCGGGAAGATTTC 126
    |||||
DB 751 AAADDAADKAAAKAKAAAAAARAAAAKAKAWAKKADDKAKAAKADADADAKKKAKA 692
    |||||
QY 67 ATTGTAGAATCGTTCGGAACAGAGGAAGAAAAACAAGCCTCCCGTCGGGAAGATTTC 126
    |||||
DB 691 AADAARAKDAXKAAAAAARAAAAAARAAAAAARAAAAAARAAAAAARAAAAAARAKK 632
    |||||
QY 127 CAACGGGTAAAAAGCGCTTGTAAATTTGGCATTTTCGGCTTGGCAACAAAATTGAAA 186
    |||||
DB 631 KKDKKAADAAAAKRAADKAKAKARAAADKAKRDAGADKKKGWKGAGGAKAAAGKXADDA 572
    |||||
QY 187 TTCTACTATCCAGGAAAAAGAGGGTTACTTTTACTTTTGGATTATTCAGAGATTTCGAA 246
    |||||
DB 571 KKKAAKTDKAKTAAKAAADAAKARTRRTAKAAATKAKAAGAKWKGWGGAKAKRTA 512
    |||||
QY 247 GAA 249
    |||||
DB 511 GAA 509
    |||||

RESULT 5
AL550892
LOCUS      AL550892 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
DEFINITION clone CSODI065Y110 5-PRIME, mRNA sequence.
ACCESSION  AL550892
VERSION     AL550892.2 GI:31272709
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens

REFERENCE   1 (bases 1 to 1201)
AUTHORS    Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE       Full-length cDNA libraries and normalization
JOURNAL     Unpublished
COMMENT     On Feb 15, 2001 this sequence version replaced gi:12889314.
            Contact: Genoscope
            Genoscope - Centre National de Séquençage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 3710.f For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CSODI065B805QPI&cluster=3710.f. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CSODI065B805QPI.
            Location/Qualifiers
                1..1201
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="CSODI065Y110"
                /tissue type="PLACENTA COT 25-NORMALIZED"
                /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                /note="1st strand cDNA was primed with a NotI-oligo(dT)
                primer. Five prime end enriched, double-strand cDNA was
                digested with Not I and cloned into the Not I and EcoRV
                sites of the pCMVSPORT 6 vector. Library was normalized."
                BASE COUNT      600 a      42 c      265 g      158 t      136 others
ORIGIN

Query Match      14.3%; Score 38.6; DB 9; Length 1201;
Best Local Similarity 40.9%; Pred. No. 8.2;
Matches 101; Conservative 21; Mismatches 125; Indels 0; Gaps 0;

QY 7 ATAAATCGTTATTAAAGATAATGTTATAGAACGAGCTTCGATGAATCAATGATTTT 66
    |||||
DB 153 AAAAAAARGTAAAMMAAAAAAARAAAAAARAAAAAARAAAAAARAAAAAARAAAAA 212
    |||||
QY 67 ATTGTAGAATCGTTCGGAACAGAGGAAGAAAAACAAGCCTCCCGTCGGGAAGATTTC 126
    |||||

```



d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelOBAC11.

FEATURES  
source Location/Qualifiers

1..859  
/organism="Drosophila melanogaster"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7227"  
/clone="BACN07C22"  
/clone\_lib="DrosBAC"  
/plasmid="pBelOBAC11"  
/notes="end : T7"

BASE COUNT 235 a 149 c 100 g 300 t 75 others  
ORIGIN

Query Match 13.8%; Score 37.2; DB 29; Length 859;  
Best Local Similarity 41.9%; Pred. No. 19;  
Matches 65; Conservative 22; Mismatches 68; Indels 0; Gaps 0;

QY 7 ATAAATCGTTATTAAAGATAAATGGTATAGCAAGAGCTTCGATGAATCAATGATTTTT 66  
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :  
Db 810 ATWKAAAGDTKATRAWTADAAGAATKAKGGAATGWKTITTRAGTTWATAAAAAATTTKAG 751  
||||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :  
QY 67 ATTGTGAATCGTGCGAACAGAGGAACAACAAAACAAAGCCCTCCCCTGCCGAAGAATTC 126  
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :  
Db 750 ATATAAAATGTAKTGTAGATARAANAATAAATAAATAATKNITTTTTTTTGTGTTT 691  
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :

QY 127 CAACGGGTAAAAAGCCCTTGTAATTTGCATTA 161  
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :  
Db 690 AAWTAGTGTAAAAAGDGTKGTTTGTGTTGAATTA 656  
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :

RESULT 10  
LOCUS BX447889 Homo sapiens T CELLS (JURKAT CELL LINE) Homo sapiens CDNA EST 22-MAY-2003  
DEFINITION clone CSODH008YD21 5-PRIME, mRNA sequence.  
ACCESSION BX447889  
VERSION 1 GI:31025988  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE Li,W.B., Gruber,C., Jessee,J. and Polayes,D.  
AUTHORS Full-length cDNA libraries and normalization  
TITLE Unpublished  
COMMENT Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600 Faraday Avenue Genoscope sequence ID : GS0AH008GB11QP1.  
Location/Qualifiers  
1..1186  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CSODH008YD21"  
/tissue\_type="T CELLS (JURKAT CELL LINE)"  
/cell\_line="JURKAT CELL LINE"  
/clone\_lib="Homo sapiens T CELLS (JURKAT CELL LINE)"  
/note="Vector: pCMVSPORT\_6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned, into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

BASE COUNT 298 a 273 c 285 g 185 t 145 others  
ORIGIN

249 CCGAGCATCACATCGTTTATTAGTTATAATCGTAAAGAGATAAAATTGTTTAAATAATAG 190

and Genevieve Payan. It has been constructed in the vector pBelobAC11.	
FEATURES	Location/Qualifiers
source	1. .1310
	/organism="Drosophila melanogaster"
	/mol_type="genomic DNA"
	/db_xref="taxon:7227"
	/clone="BACN15124"
	/clone_lib="DrosBAC"
	/plasmid="pBelobAC11"
	/note="end : Sp8"
BASE COUNT	330 a 140 C 219 G 396 t 225 others
ORIGIN	
Query Match	13.8%; Score 36.6; DB 29; Length 1310;
Best Local Similarity	34.8%; Pred. No. 29;
Matches	63; Conservative 39; Mismatches 79; Indels 0; Gaps 0;
QY	62 TTTTATTTCAGAAATCGGTGGGAACAGAGGGGAAGAAAACAAGCCTCCCGTGCGAAG 1215
Db	1036 TTWTTTATTATAAAAAAAMWAAAAAAMWAAAAAAMWAAAAAAMWAAAAAAGTTTWTDKDRKKGGK 1095
QY	122 AATTCACGGGTAAAAAGCGCTTCGTAAATGGCATTATTCGGGCTTGGCAACAAAAT 181
Db	1096 ATTAAATATRAWAAAAAAMWGGTGGGGAAAAAGKKTDRREKGTDKWAKGAAAWRAG 1155
QY	182 TGAATTTCTACTATCCAGGAAACGAAAGGTTACTTTTACTTTCGATTATGCAGAGAT 241
Db	1156 WTTTWWTTKDTKTTTGAATTTWADRAWAANDRDKKRWRTGTTTTTTATTGGGGGAT 1215
QY	242 T 242
Db	1216 T 1216
RESULT 15	
AZ674925/c	
LOCUS	885 bp DNA linear GSS 14-DEC-2000
DEFINITION	ENTUM21FR Entamoeba histolytica Sheared DNA Entamoeba histolytica genomic, genomic survey sequence.
ACCESSION	AZ674925.1 GI:11812071
VERSION	GSS.
KEYWORDS	Entamoeba histolytica
SOURCE	Entamoeba histolytica
ORGANISM	Eukaryota; Entamoebidae; Entamoeba.
REFERENCE	1 (bases 1 to 885)
AUTHORS	Lofus, B., Van Aken, S. and Fraser, C.
TITLE	Determination of clone end sequences from Entamoeba histolytica
JOURNAL	HM1:IMSS sheared DNA library
COMMENT	Unpublished Contact: Brendan J Loftus Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0208 Fax: 301 838 3543 Email: bjloftus@tigr.org Clones are derived from the Entamoeba histolytica HM1:IMSS sheared DNA library Seq primer: M13-Reverse Class: shotgun High quality sequence start: 19 High quality sequence stop: 853.
FEATURES	Location/Qualifiers
source	1. .885
	/organism="Entamoeba histolytica"
	/mol_type="genomic DNA"
	/strain="HM1:IMSS"
	/db_xref="taxon:5759"
	/clone_lib="Entamoeba histolytica Sheared DNA"
	/note="Vector: pHO31; Site 1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD."

Genomic DNA isolated from broth cultures of *E. histolytica* using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) *Entamoeba histolytica*: a method for isolate identification. *Exp. Parasitol.* 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The *vs-i* method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In *Genome Sequencing: A Practical Approach*, eds. M. Vaudin and B. Barell, Oxford University Press, 1993).

BASE COUNT 189 a 120 c 135 g 441 t  
ORIGIN  
Query Match 13.5%; Score 36.4; DB 28; Length 885;  
Best Local Similarity 50.6%; Pred. No. 31;  
Matches 88; Conservative 0; Mismatches 86; Indels 0; Gaps 0;  
QY 94 AAGAAACACAGCCTCCCGTGGGAGCAATTCCACGGGTAAATAAGGCTTGGTAAT 153  
Db 195 AGGAAACCAATTCAATACACCGTTTATTCCTCAAAAACAGAGAAACAAAAAT 136  
QY 154 TGGCATTATTCGGGTTGGCAACAAAATTGAAATTTCTACTATCCAGGGAAACGAAAGG 213  
Db 135 TAGTATTAAATTGAAGATGACCAATAAATAAAGAGATATAACACAGAGAGATGAAGAGG 76  
QY 214 TTACTTTTACTTTGGAATTATCCAGAGATTGGAAGACGAAGGGTATTATTATCAAC 267  
Db 75 TTCAAGAGAAAGTGACTAATGACGACCAATCAAGAAAGACTGAAGAAACATCATC 22

Search completed: November 15, 2003, 08:02:28  
Job time : 1798.52 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 15, 2003, 00:06:39 ; Search time 41.8989 Seconds  
(without alignments)  
2833.774 Million cell updates/sec

Title: US-09-928-457-77

Perfect score: 269

Sequence: 1 CGGAGCTAATAATCGTTATT.....GCAAGGTTATTATCAACG 269

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.\*

- 1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*
- 2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*
- 3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*
- 4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*
- 5: /cgn2\_6/ptodata/2/ina/PTCUS\_COMB.seq.\*
- 6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
C 1	35	13.0	7218	1	US-08-232-463-14
C 2	32.4	12.0	1664976	4	US-08-916-421B-1
C 3	32	11.9	6243	2	US-08-056-075-1
C 4	31.8	11.8	435	4	US-09-134-001C-123
C 5	31.6	11.7	1167	4	US-09-107-532A-2727
C 6	31	11.5	1713	4	US-09-134-001C-135
C 7	31	11.5	1920	4	US-09-134-001C-2824
C 8	31	11.5	1230025	4	US-09-198-452A-1
C 9	30.2	11.2	1031	1	US-08-181-271A-106
C 10	30.2	11.2	1031	1	US-08-449-315-106
C 11	30.2	11.2	1031	1	US-08-444-803-106
C 12	30.2	11.2	1031	1	US-08-449-043-106
C 13	30.2	11.2	1031	1	US-08-456-265A-106
C 14	30.2	11.2	1031	1	US-08-458-416-106
C 15	30.2	11.2	1031	1	US-08-455-244-106
C 16	30.2	11.2	1031	1	US-08-454-876-106
C 17	30.2	11.2	1031	2	US-08-457-364-106
C 18	30.2	11.2	1031	2	US-08-456-262-106
C 19	30.2	11.2	1031	2	US-08-456-340-106
C 20	30.2	11.2	1031	2	US-08-455-736-106
C 21	30.2	11.2	1031	2	US-08-971-217-106
C 22	30.2	11.2	1031	3	US-09-350-600-106
C 23	29.8	11.1	1497	4	US-09-220-132-94
C 24	29.8	11.1	5656	1	US-08-425-061-1
C 25	29.8	11.1	5656	1	US-08-825-886-1
C 26	29.8	11.1	5656	4	US-08-989-890-1
C 27	29.8	11.1	5689	1	US-08-425-061-3

28	29.8	11.1	5689	1	US-08-825-886-3	Sequence 3, Appli
29	29.8	11.1	5689	4	US-08-989-890-3	Sequence 3, Appli
30	29.8	11.1	5709	1	US-08-425-061-2	Sequence 2, Appli
31	29.8	11.1	5709	1	US-08-425-061-7	Sequence 7, Appli
32	29.8	11.1	5709	1	US-08-425-061-8	Sequence 8, Appli
33	29.8	11.1	5709	1	US-08-425-061-9	Sequence 9, Appli
34	29.8	11.1	5709	1	US-08-825-886-2	Sequence 2, Appli
35	29.8	11.1	5709	1	US-08-825-886-7	Sequence 7, Appli
36	29.8	11.1	5709	1	US-08-825-886-8	Sequence 8, Appli
37	29.8	11.1	5709	1	US-08-825-886-9	Sequence 9, Appli
38	29.8	11.1	5709	4	US-08-989-890-2	Sequence 2, Appli
39	29.8	11.1	5709	4	US-08-989-890-7	Sequence 7, Appli
40	29.8	11.1	5709	4	US-08-989-890-8	Sequence 8, Appli
41	29.8	11.1	5709	4	US-08-989-890-9	Sequence 9, Appli
42	29.8	11.1	5710	1	US-08-425-061-6	Sequence 6, Appli
43	29.8	11.1	5710	1	US-08-825-886-6	Sequence 6, Appli
44	29.8	11.1	5710	4	US-08-989-890-6	Sequence 6, Appli
45	29.8	11.1	5711	1	US-08-425-061-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1  
US-08-232-463-14/c  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFLINGER, F.  
; APPLICANT: FALKNER, F. G.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 1800 Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22313-0299  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/232,463  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/935,313  
; FILING DATE:  
; APPLICATION NUMBER: EP 91 114 300.6  
; FILING DATE: 26-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 30472/114 INMU  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703)836-9300  
; TELEFAX: (703)683-4109  
; TELEX: 899149  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7218 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; CLONE: PTZ9PT-F15  
US-08-232-463-14  
Query Match 13.0%; Score 35; DB 1; Length 7218;

```
Best Local Similarity 4.6%; Pred. No. 0.36;
Matches 11; Conservative 134; Mismatches 94; Indels 0; Gaps 0;
Qy 18 ATTAAGATAATGTTATGAGGAGGCTTCGATGAAATCAATGATTTTATTGAGAATC 77
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1451 ATAGAAGAAATTTGTTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1392
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 78 GTCGGAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 137
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1391 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1332
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 138 AAAGCGCTTGGTAAATGGCATTTTCGGCTTGGCAACAAATGAAATTTCTACTATC 197
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1331 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1272
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 198 CAGGAAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 256
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1271 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1213
    : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 2
US-08-916-421B-1
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
; APPLICANT: Bult et al.
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococcus
; Patent No. 6503729
; TITLE OF INVENTION: jannaschii
; FILE REFERENCE: PB275
; CURRENT APPLICATION NUMBER: US/08/916,421B
; CURRENT FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/024,428
; PRIOR FILING DATE: 1996-08-22
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1664976
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (28222)..(28222)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (28257)..(28258)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84773)..(84773)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84808)..(84808)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84812)..(84812)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98120)..(98120)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98159)..(98159)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98239)..(98239)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98266)..(98266)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98343)..(98343)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (103998)..(103998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (148948)..(148948)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (163385)..(163385)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (191989)..(191989)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (191995)..(191995)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (231980)..(231980)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234187)..(234187)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
```



;; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

;; FILE REFERENCE: GTC-007  
;; CURRENT APPLICATION NUMBER: US/09/134,001C  
;; CURRENT FILING DATE: 1998-08-13  
;; PRIOR APPLICATION NUMBER: US 60/064,964  
;; PRIOR FILING DATE: 1997-11-08  
;; PRIOR APPLICATION NUMBER: US 60/055,779  
;; PRIOR FILING DATE: 1997-08-14  
;; NUMBER OF SEQ ID NOS: 5674  
;; SEQ ID NO 123  
;; LENGTH: 435  
;; TYPE: DNA

;; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-123

Query Match 11.8%; Score 31.8; DB 4; Length 435;  
Best Local Similarity 53.7%; Pred. No. 1.4;  
Matches 66; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY 9 AAAATCGTTATTAAAGATAATGGTATAGCAAGAGCTTCGATGGAATCAATGATTTTAT 68  
DB 192 AGATTACTATTAGATATGAGCTTTGTTGGTGATGCTATGATGAAGCATGAAACCTAC 251  
QY 69 TTGAGAAATCGTTCGGAACAGAGGAAGAAACAAAGCCTCCCGTCGGAAGAAATCCA 128  
DB 252 AAGAGTAATATGAAATAGTATAGTAAATAAATCACTTTCTTCATGCACATTTATCCA 311  
QY 129 ACG 131  
DB 312 AAG 314

## RESULT 5

US-09-107-532A-2727/c  
;; Sequence 2727, Application US/09107532A  
;; Patent No. 6583275

## GENERAL INFORMATION:

;; APPLICANT: Lynn A Doucette-Stamm and David Bush  
;; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS

;; NUMBER OF SEQUENCES: 7310

## CORRESPONDENCE ADDRESS:

;; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
;; STREET: 100 Beaver Street  
;; CITY: Waltham  
;; STATE: Massachusetts  
;; COUNTRY: USA  
;; ZIP: 02354

## COMPUTER READABLE FORM:

;; MEDIUM TYPE: CD-ROM ISO9660  
;; COMPUTER: PC  
;; OPERATING SYSTEM: <Unknown>  
;; SOFTWARE: ASCII

;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/107,532A

;; FILING DATE: 30-Jun-1998

## PRIOR APPLICATION DATA:

;; APPLICATION DATA:  
;; APPLICATION NUMBER: 60/085,598  
;; FILING DATE: 14 May 1998  
;; APPLICATION NUMBER: 60/051571  
;; FILING DATE: July 2, 1997

## ATTORNEY/AGENT INFORMATION:

;; NAME: Arinello, Pamela Deneke  
;; REGISTRATION NUMBER: 40,489  
;; REFERENCE/DOCKET NUMBER: GTC-012  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (781)893-5007  
;; TELEFAX: (781)893-6277

;; INFORMATION FOR SEQ ID NO: 2727:

## SEQUENCE CHARACTERISTICS:

;; LENGTH: 1167 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: double

;; TOPOLOGY: circular  
;; MOLECULE TYPE: DNA (genomic)  
;; HYPOTHETICAL: NO  
;; ANTI-SENSE: NO  
;; ORIGINAL SOURCE:  
;; ORGANISM: Enterococcus faecium  
;; FEATURE:  
;; NAME/KEY: misc\_feature  
;; LOCATION: (B) LOCATION 1...1167  
;; SEQUENCE DESCRIPTION: SEQ ID NO: 2727:  
US-09-107-532A-2727

Query Match 11.7%; Score 31.6; DB 4; Length 1167;  
Best Local Similarity 51.4%; Pred. No. 2.2;  
Matches 73; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 10 AAATCGTTATTAAAGATAATGGTATAGCAAGAGCTTCGATGGAATCAATGATTTTAT 69  
DB 1064 AAATCAGTGTAGAAATAAAGTTTATACAAACTTTTCGCTGATCAAGTTGACGAGATTTAAT 1005

QY 70 TCAGAAATCGTTCGGAACAGAGGGAAGAAACAAAGCCTCCCGTCGGAAGAAATCCA 129  
DB 1004 TCACCAACAATCGGAATATGAAGATCAATCCAATCTTCACTTGTTCGCCCAATCAT 945

QY 130 CGGTAATAAAGGCGCTTGTTAA 151  
DB 944 CCAGCAAAAAGCTCATCGCA 923

## RESULT 6

US-09-134-001C-135  
;; Sequence 135, Application US/09134001C  
;; Patent No. 6380370

## GENERAL INFORMATION:

;; APPLICANT: Lynn Doucette-Stamm et al  
;; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

;; FILE REFERENCE: GTC-007

;; CURRENT APPLICATION NUMBER: US/09/134,001C

;; CURRENT FILING DATE: 1998-08-13

;; PRIOR APPLICATION NUMBER: US 60/064,964

;; PRIOR FILING DATE: 1997-11-08

;; PRIOR APPLICATION NUMBER: US 60/055,779

;; PRIOR FILING DATE: 1997-08-14

;; NUMBER OF SEQ ID NOS: 5674

;; SEQ ID NO 135

;; LENGTH: 1713

;; TYPE: DNA

;; ORGANISM: Staphylococcus epidermidis

US-09-134-001C-135

Query Match 11.5%; Score 31; DB 4; Length 1713;

Best Local Similarity 51.9%; Pred. No. 3.7;

Matches 70; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 97 AAAACAAGCTCCCGTCGGAAGAAATCCAAAGGTAATAAAGGCGCTTCGTAATGG 156  
DB 1284 AAAACAACAAAAAATAATGAAGAAATTAATCTGGGAATATTTTAATCCCAATTA 1343  
QY 157 CATTATTCGGGCTTGGCAACAAAATTGAAATTTCTACTATCCAGGGAACCAAGGGTTA 216  
DB 1344 CAAAAGCTTTCAGAACCCAGAACAAAAAATATCTACAGTCAAGAAAAAATTGATGGA 1403

QY 217 CTTTACITTTGGATT 231  
DB 1404 ACCTGTTTITGGATT 1418

## RESULT 7

US-09-134-001C-2824  
;; Sequence 2824, Application US/09134001C  
;; Patent No. 6380370  
;; GENERAL INFORMATION:

```
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 2824
; LENGTH: 1920
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
; US-09-134-001C-2824

Query Match      11.5%; Score 31; DB 4; Length 1920;
Best Local Similarity 46.9%; Pred. No. 3.8;
Matches 97; Conservative 0; Mismatches 110; Indels 0; Gaps 0;

QY 20 TAAAGATATGTTATAGACAGCTTCGATGAATCAATGATTTTATTTGAGATCGG 79
Db 1341 TGATTTGATGCTGTAGACCGACACTTATTTTAAATGATGAAGATTATCAAAAAATTAG 1400
QY 80 TCGGAACAGAGGGAAGAAACAAAGCTCCCGTGGCGGAAGATTCACACGGGTAAAA 139
Db 1401 AGAACAATATTAAGAAAGAAATATTTTCCCAATATGTTTGAATCTTAAAAATAAAA 1460
QY 140 AGGCTTGGTAAATGGCATTATTCGGCTTGGCAACAAAATTGAAATTTCTACTATCCA 199
Db 1461 TGATTTACCAATATGAAAAATAGTTAGTAGTATCAATCAAGATATAGAAACAAGAAG 1520
QY 200 GGAACACGAAGGGTACTTTACTTT 226
Db 1521 TGACCGCAAGTGAATTTCTAGTTT 1547

RESULT 8
US-09-198-452A-1/c
; Sequence 1, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Grifais, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
; TITLE OF INVENTION: and treatment of infection
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198,452A
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 1
; LENGTH: 1230025
; TYPE: DNA
; ORGANISM: Chlamydia pneumoniae
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..<15000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (15001)..<30000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (30001)..<45000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (45001)..<60000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (60001)..<75000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (75001)..<90000)
; OTHER INFORMATION: n=a or c or g or t

; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (90001)..<105000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (105001)..<120000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (120001)..<135000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (135001)..<150000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (150001)..<165000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (165001)..<180000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (180001)..<195000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (195001)..<210000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (210001)..<225000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (225001)..<240000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (240001)..<255000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (255001)..<270000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (270001)..<285000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (285001)..<300000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (300001)..<315000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (315001)..<330000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (330001)..<345000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (345001)..<360000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (360001)..<375000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (375001)..<390000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (390001)..<405000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (405001)..<420000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (420001)..<435000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (435001)..<450000)
; OTHER INFORMATION: n=a or c or g or t
```

NAME/KEY: misc feature  
LOCATION: (450001)..(465000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc feature  
LOCATION: (465001)..(480000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc feature  
LOCATION: (480001)..(495000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc feature  
LOCATION: (495001)..(510000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc feature  
LOCATION: (510001)..(525000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc feature  
LOCATION: (525001)..(540000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc feature  
LOCATION: (540001)..(555000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc feature  
LOCATION: (555001)..(570000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc feature  
LOCATION: (570001)..(585000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc feature  
LOCATION: (585001)..(600000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc feature  
LOCATION: (600001)..(615000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc feature  
LOCATION: (615001)..(630000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc feature  
LOCATION: (630001)..(645000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc feature  
LOCATION: (645001)..(660000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc feature  
LOCATION: (660001)..(675000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc feature  
LOCATION: (675001)..(690000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc feature  
LOCATION: (690001)..(705000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc feature  
LOCATION: (705001)..(720000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc feature  
LOCATION: (720001)..(735000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc feature  
LOCATION: (735001)..(750000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc feature  
LOCATION: (750001)..(765000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc feature  
LOCATION: (765001)..(780000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc feature  
LOCATION: (780001)..(795000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc feature  
LOCATION: (795001)..(810000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc feature

LOCATION: (810001)..(825000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc feature  
LOCATION: (825001)..(840000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc feature  
LOCATION: (840001)..(855000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc feature  
LOCATION: (855001)..(870000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc feature  
LOCATION: (870001)..(885000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc feature  
LOCATION: (885001)..(900000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc feature  
LOCATION: (900001)..(915000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc feature

Query Match 11.5%; Score 31; DB 4; Length 1230025;  
Best Local Similarity 64.8%; Pred. No. 25;  
Matches 46; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 175 ACAAAATTGAAATTTCTACTATCCAGGAAACGAAAGGGTTACTTTTACTTTGGATTATG 234  
Db 797606 ACCAACTACAAATCTATTATCCAGGAAACGAAAGGGTTACTTTTACTTTGGATTATG 797547  
QY 235 CAGAGATTGCA 245  
Db 797546 AAGATATTGGA 797536

RESULT 9

US-08-181-271A-106/c  
; Sequence 106, Application US/08181271A  
; Patent No. 5614395  
; GENERAL INFORMATION:  
; APPLICANT: Ryals, John A.  
; APPLICANT: Alexander, Danny C.  
; APPLICANT: Beck, James J.  
; APPLICANT: Duesing, John H.  
; APPLICANT: Friedrich, Leslie B.  
; APPLICANT: Goodman, Robert M.  
; APPLICANT: Harms, Christian  
; APPLICANT: Meins, Jr., Frederick  
; APPLICANT: Montoya, Alice  
; APPLICANT: Moyer, Mary B.  
; APPLICANT: Neuhaus, Jean-Marc  
; APPLICANT: Payne, George B.  
; APPLICANT: Sperison, Christoph  
; APPLICANT: Stinson, Jeffrey R.  
; APPLICANT: Uknes, Scott J.  
; APPLICANT: Ward, Eric R.  
; APPLICANT: Williams, Shericka C.  
; TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC  
; NUMBER OF SEQUENCES: 106  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CIBA-GEIGY Corporation  
; STREET: 7 Skyline Drive  
; CITY: Hawthorne  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10532

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:

```

; APPLICATION NUMBER: US/08/181,271A
; FILING DATE: 13-JAN-94
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/093,301
; FILING DATE: 16-JUL-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/937,197
; FILING DATE: 6-NOV-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/678,378
; FILING DATE: 1-APR-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/305,566
; FILING DATE: 6-FEB-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/165,667
; FILING DATE: 8-MAR-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/042,847
; FILING DATE: 6-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/632,441
; FILING DATE: 21-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/425,504
; FILING DATE: 20-OCT-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/848,506
; FILING DATE: 6-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/768,122
; FILING DATE: 27-SEP-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/580,431
; FILING DATE: 7-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/368,672
; FILING DATE: 20-JUN-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/329,018
; FILING DATE: 24-MAR-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/045,957
; FILING DATE: 12-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919)541-8614
; TELEFAX: (919)541-8689
; INFORMATION FOR SEQ ID NO: 106:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1031 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-181-271A-106

Query Match 11.2%; Score 30.2; DB 1; Length 1031;
Best Local Similarity 45.3%; Pred. No. 5.5;
Matches 110; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

QY - 21 AAAGATAATGGTATAGGAACGAGCTTCGATGAATCAATGATTTTATTGAGATCGGT 80
DB 256 AAACATTTGCTTTGTTTGGTTTTCACATCGATCGAGTTGCTGACACGAGGTT 197
QY 81 CGGAACAGAGGGAAGAAAACAGCCCTCCCGTGGGAAGAAATCCCAACGGGTAAAAA 140
DB 196 CAGTGTACCGCGCGAGTAATCACCAGATGACCCATGAACAAATGGAGACCTGGATTAA 137

QY 141 GGCTTTGGTAATGGCATTTATTCGGGCTTGGCAACAAAATTCGAAATTTCTACTATCCAG 200
DB 136 TACTTTTGGAAATTCGCAAAATCTCTCCCTTGACCGCAACATCTGATTTGCTTCAAGCAC 77
QY 201 GGAACGAAAGGCTTACTTTTACTTTTGGATTATGCAGAGATTCGAAGACGAGGATTT 260
DB 76 GCAACAGAAAAGATTCCTCATTTTCTGAACAAGAAAGCAAGAGATTATGATGGCTTAT 17
QY 261 TAT 263
DB 16 TAT 14

RESULT 10
US-08-449-315-106/c
; Sequence 106, Application US/08449315
; Patent No. 5650505
; GENERAL INFORMATION:
; APPLICANT: Rvals, John A.
; APPLICANT: Alexander, Danny C.
; APPLICANT: Beck, James J.
; APPLICANT: Duesing, John H.
; APPLICANT: Friedrich, Leslie B.
; APPLICANT: Goodman, Robert M.
; APPLICANT: Harms, Christian
; APPLICANT: Meins, Jr. Frederick
; APPLICANT: Montoya, Alice
; APPLICANT: Moyer, Mary B.
; APPLICANT: Neuhaus, Jean-Marc
; APPLICANT: Payne, George B.
; APPLICANT: Sperison, Christoph
; APPLICANT: Stinson, Jeffrey R.
; APPLICANT: Uknes, Scott J.
; APPLICANT: Ward, Eric R.
; APPLICANT: Williams, Shericca C.
; TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
; TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF
; NUMBER OF SEQUENCES: 106
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: New York
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/449,315
; FILING DATE: 24-MAY-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/181,271
; FILING DATE: 13-JAN-94
; APPLICATION NUMBER: US 08/093,301
; FILING DATE: 16-JUL-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/937,197
; FILING DATE: 6-NOV-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/678,378
; FILING DATE: 1-APR-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/305,566
; FILING DATE: 6-FEB-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/165,667
; FILING DATE: 8-MAR-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/042,847
```

;; FILING DATE: 6-APR-1993  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/632,441  
;; FILING DATE: 21-DEC-1990  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/425,504  
;; FILING DATE: 20-OCT-1989  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/948,506  
;; FILING DATE: 6-MAR-1992  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/768,122  
;; FILING DATE: 27-SEP-1991  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/580,431  
;; FILING DATE: 7-SEP-1990  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/368,672  
;; FILING DATE: 20-JUN-1989  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/329,018  
;; FILING DATE: 24-MAR-1989  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/045,957  
;; FILING DATE: 12-APR-1993  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Elmet, James Scott  
;; REGISTRATION NUMBER: 36,129  
;; REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (919)541-8614  
;; TELEFAX: (919)541-8699  
;; INFORMATION FOR SEQ ID NO: 106:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1031 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: DNA (genomic)  
US-08-449-315-106

Query Match 11.2%; Score 30.2; DB 1; Length 1031;  
Best Local Similarity 45.3%; Pred. No. 5.5;  
Matches 110; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

QY 21 AAAGATATGGTATAGGAACGAGCTTCGATGAATCAATGATTTTATTTGAGAAATCGGT 80  
DB 256 AAACATATGCTTTGTTTGGTTTTCACATGACATCGAGTTGCTGAAGACGAGGT 197

QY 81 CGAACAAGAGGAG 140  
DB 196 CAGTGACGCGGACGATTAACCGATGACCCATGGAACAAATGGAGACCTGGATTAAG 137

QY 141 GCCTTGTGAATTCGATTTTCGGCTTCGCAACAAATGAAATTTCTACTATCCAG 200  
DB 136 TACTTTTGGAAATTTGCAAAATCTCTCTTACCGCAACATCTGATTCCTTCAACGAC 77

QY 201 GAAACAGAAAGAGGTTACTTTTACTTTGATTTGATTCAGAGATTCGAAGAGAGAGGTTAT 260  
DB 76 GCAACAGAAAGATTCCTCCCATTTATCTGAACAGAGCAAGATGATTAATGATTCAT 17

QY 261 TAT 263  
DB 16 TAT 14

RESULT 11  
US-08-444-803-106/C  
; Sequence 106, Application US/08444803  
; Patent No. 5654414  
; GENERAL INFORMATION:  
; APPLICANT: Ryals, John A.  
; APPLICANT: Alexander, Danny C.

;; APPLICANT: Beck, James J.  
;; APPLICANT: Duesing, John H.  
;; APPLICANT: Friedrich, Leslie B.  
;; APPLICANT: Goodman, Robert M.  
;; APPLICANT: Harms, Christian  
;; APPLICANT: Meins, Jr., Frederick  
;; APPLICANT: Montoya, Alice  
;; APPLICANT: Moyer, Mary B.  
;; APPLICANT: Neuhaus, Jean-Marc  
;; APPLICANT: Payne, George B.  
;; APPLICANT: Sperison, Christoph  
;; APPLICANT: Stinson, Jeffrey R.  
;; APPLICANT: Uknes, Scott J.  
;; APPLICANT: Ward, Eric R.  
;; APPLICANT: Williams, Shericca C.  
;; TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC  
;; TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF  
;; NUMBER OF SEQUENCES: 106  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: CIBA-GEIGY Corporation  
;; STREET: 7 Skyline Drive  
;; CITY: Hawthorne  
;; STATE: New York  
;; COUNTRY: USA  
;; ZIP: 10532  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patent in Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/444,803  
;; FILING DATE: 19-MAY-1995  
;; CLASSIFICATION: 536  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/181,271  
;; FILING DATE: 13-JAN-94  
;; APPLICATION NUMBER: US 08/093,301  
;; FILING DATE: 16-JUL-1993  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/937,197  
;; FILING DATE: 6-NOV-1992  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/678,378  
;; FILING DATE: 1-APR-1991  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/305,566  
;; FILING DATE: 6-FEB-1989  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/185,667  
;; FILING DATE: 8-MAR-1988  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/042,847  
;; FILING DATE: 6-APR-1993  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/632,441  
;; FILING DATE: 21-DEC-1990  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/425,504  
;; FILING DATE: 20-OCT-1989  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/768,122  
;; FILING DATE: 27-SEP-1991  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/580,431  
;; FILING DATE: 7-SEP-1990  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/368,672  
;; FILING DATE: 20-JUN-1989  
;; PRIOR APPLICATION DATA:



```

/ APPLICATION NUMBER: US 07/329,018
/ FILING DATE: 24-MAR-1989
/ PRIORITY APPLICATION DATA:
/ APPLICATION NUMBER: US 08/045,957
/ FILING DATE: 12-APR-1993
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Elmer, James Scott
/ REGISTRATION NUMBER: 36,129
/ REFERENCE/DOCKET NUMBER: S-19825/Pl/CGC 1727
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (919)541-8614
/ TELEFAX: (919)541-8689
/ INFORMATION FOR SEQ ID NO: 106:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1031 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: DNA (genomic)
/ US-08-444-803-106

Query Match 11.2%; Score 30.2; DB 1; Length 1031;
Best Local Similarity 45.3%; Pred. No. 5.5;
Matches 110; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

QY 21 AAAGATAATGTTATAGGAACGAGCTTCGATGAAATCAATGATTTTTTATTGAGAATCGGT 80
DB 256 AAACACATTGCTCTTTGTTTTCACACTGACATCGAGTTGCTGAGACGCAAGGTT 197
QY 81 CGGAACAGAGGAGGAGAAAAACAAGCCTCCCGTCGGGAGAAATTCGAACGGGTAAAAA 140
DB 196 CAGTGTACCGGCGAGTAAATCACCGATGACCCATGGAACAAATGGAGAGCTCGAATTAAG 137
QY 141 GGCCTTGTAAATTTGGCATTTATCGGGCTTGGCAACAAATTTGAAATTTTACTATCCAG 200
DB 136 TACTTTTGGAAATTTGCAAAATCTCTCTTACCGCAACATCTGATTGCTTCAACAGCAC 77
QY 201 GGAACGAAAGGGTTACTTTTACTTTGGATTATGCAGAGATTCGAAGAAGCAAGGGTATT 260
DB 76 CCAACAGAAAGATTCCCCATTATTCTGCAACAGAGCAAGCAAGATTATGAATGGCTTAT 17
QY 261 TAT 263
DB 16 TAT 14

RESULT 12
US-08-449-043-106/C
/ Sequence 106, Application US/08449043
/ Patent No. 5689044
/ GENERAL INFORMATION:
/ APPLICANT: Ryals, John A.
/ APPLICANT: Alexander, Danny C.
/ APPLICANT: Beck, James J.
/ APPLICANT: Duesing, John H.
/ APPLICANT: Friedrich, Leslie B.
/ APPLICANT: Goodman, Robert M.
/ APPLICANT: Harms, Christian
/ APPLICANT: Meins, Jr., Frederick
/ APPLICANT: Montoya, Alice
/ APPLICANT: Moyer, Mary B.
/ APPLICANT: Neuhaus, Jean-Marc
/ APPLICANT: Payne, George B.
/ APPLICANT: Sperison, Christoph
/ APPLICANT: Stinson, Jeffrey R.
/ APPLICANT: Uknes, Scott J.
/ APPLICANT: Ward, Eric R.
/ APPLICANT: Williams, Shericca C.
/ TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
/ NUMBER OF SEQUENCES: 106
/ CORRESPONDENCE ADDRESS:
/ ADDRESS: CIBA-GEIGY Corporation
/ STREET: 7 Skyline Drive
/ CITY: Hawthorne
/ STATE: New York
/ COUNTRY: USA
/ ZIP: 10532
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/449,043
/ FILING DATE: 24-MAY-1995
/ CLASSIFICATION: 800
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 08/181,271
/ FILING DATE: 13-JAN-94
/ APPLICATION NUMBER: US 08/093,301
/ FILING DATE: 16-JUL-1993
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 07/937,197
/ FILING DATE: 6-NOV-1992
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 07/678,378
/ FILING DATE: 1-APR-1991
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 07/305,566
/ FILING DATE: 6-FEB-1989
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 07/165,667
/ FILING DATE: 8-MAR-1988
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 08/042,847
/ FILING DATE: 6-APR-1993
/ APPLICATION NUMBER: US 07/632,441
/ FILING DATE: 21-DEC-1990
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 07/425,504
/ FILING DATE: 20-OCT-1989
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 07/848,506
/ FILING DATE: 6-MAR-1992
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 07/768,122
/ FILING DATE: 27-SEP-1991
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 07/580,431
/ FILING DATE: 7-SEP-1990
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 07/368,672
/ FILING DATE: 20-JUN-1989
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 07/329,018
/ FILING DATE: 24-MAR-1989
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 08/045,957
/ FILING DATE: 13-APR-1993
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Elmer, James Scott
/ REGISTRATION NUMBER: 36,129
/ REFERENCE/DOCKET NUMBER: S-19825/Pl/CGC 1727
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (919)541-8614
/ TELEFAX: (919)541-8689
/ INFORMATION FOR SEQ ID NO: 106:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1031 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: DNA (genomic)
/ US-08-449-043-106

```

Query Match 11.2%; Score 30.2; DB 1; Length 1031;  
Best Local Similarity 45.3%; Pred. No. 5.5;  
Matches 110; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

21 AAAGATAATGGTATAGGAACGAGCTTCGATGAAATCAATGATTTTATTGAGAAATCGGT 80  
|||  
256 AACACATTTGCTTTGTTTTCACATGACATCGAGTTGCTGAACGAGGTT 197  
|||  
81 CGAACAAGAGGAGGAGAAACAAACGCTCCCGTCCGGAAGAAATCCAAACGGGTAAAAA 140  
|||  
196 CAGTGTACCGCGAGCTTAATCACCAGTACCCATGACCAATTCGAGACCTGGATTAA 137  
|||  
141 GGCCTTGGTAAATTCGCATTATTCGGCTTGGCAACAAATTTGAATTTCTACTATCCAG 200  
|||  
136 TACTTTTGGAAATTTGCCAATCTCTCTTTCACCCACATCTGTTCTTCACAGCAC 77  
|||  
201 GGAACGAAAGGTTACTTTTACTTTTGGATTATCGAGAGATTCGAGAGCAAGGTTAT 260  
|||  
76 GCAACAGAAAGATTCCTCCATTTCTGCAACAGAAAGCAAGATTATGAATGCTTAT 17  
|||  
261 TAT 263  
|||  
16 TAT 14  
|||

RESULT 13  
US-08-456-265A-106/c  
; Sequence 106 Application US/08456265A  
; Patent No. 5767369  
; GENERAL INFORMATION:  
; APPLICANT: Alexander, Danny C.  
; APPLICANT: Ryals, John A.  
; APPLICANT: Goodman, Robert M.  
; APPLICANT: Stinson, Jeffrey R.  
; TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC  
; TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF  
; NUMBER OF SEQUENCES: 111  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: CIBA-GEIGY Corporation  
; STREET: 520 White Plains Road, P.O. Box 2005  
; CITY: Tarrytown  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10591  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/456,265A  
; FILING DATE: 31-MAY-95  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/181,271  
; FILING DATE: 13-JAN-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/678,378  
; FILING DATE: 1-APR-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/305,566  
; FILING DATE: 6-FEB-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/165,667  
; FILING DATE: 8-MAR-1988  
; PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/042,847  
FILING DATE: 6-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/632,441  
FILING DATE: 21-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/425,504  
FILING DATE: 20-OCT-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/848,506  
FILING DATE: 6-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/768,122  
FILING DATE: 27-SEP-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/580,431  
FILING DATE: 7-SEP-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/368,672  
FILING DATE: 20-JUN-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/329,018  
FILING DATE: 24-MAR-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/045,957  
FILING DATE: 12-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Meigs, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: S-19825/PL/CGC 1727/DIV10  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919)541-8587  
TELEFAX: (919)541-8689  
INFORMATION FOR SEQ ID NO: 106:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1031 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-456-265A-106

Query Match 11.2%; Score 30.2; DB 1; Length 1031;  
Best Local Similarity 45.3%; Pred. No. 5.5;  
Matches 110; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

21 AAAGATAATGGTATAGGAACGAGCTTCGATGAAATCAATGATTTTATTGAGAAATCGGT 80  
|||  
256 AACACATTTGCTTTGTTTTCACATGACATCGAGTTGCTGAACGAGGTT 197  
|||  
81 CGAACAAGAGGAGGAGAAACAAACGCTCCCGTCCGGAAGAAATCCAAACGGGTAAAAA 140  
|||  
196 CAGTGTACCGCGAGCTTAATCACCAGTACCCATGACCAATTCGAGACCTGGATTAA 137  
|||  
141 GGCCTTGGTAAATTCGCATTATTCGGCTTGGCAACAAATTTGAATTTCTACTATCCAG 200  
|||  
136 TACTTTTGGAAATTTGCCAATCTCTCTTTCACCCACATCTGTTCTTCACAGCAC 77  
|||  
201 GGAACGAAAGGTTACTTTTACTTTTGGATTATCGAGAGATTCGAGAGCAAGGTTAT 260  
|||  
76 GCAACAGAAAGATTCCTCCATTTCTGCAACAGAAAGCAAGATTATGAATGCTTAT 17  
|||  
261 TAT 263  
|||  
16 TAT 14  
|||

RESULT 14  
US-08-456-416-106/c  
; Sequence 106 Application US/08455416  
; Patent No. 5777200  
; GENERAL INFORMATION:  
; APPLICANT: Ryals, John A.

APPLICANT: Alexander, Danny C.  
APPLICANT: Beck, James J.  
APPLICANT: Duesing, John H.  
APPLICANT: Friedrich, Leslie B.  
APPLICANT: Goodman, Robert M.  
APPLICANT: Harms, Christian  
APPLICANT: Meins, Jr., Frederick  
APPLICANT: Montoya, Alice  
APPLICANT: Moyer, Mary B.  
APPLICANT: Neuhaus, Jean-Marc  
APPLICANT: Payne, George B.  
APPLICANT: Sperison, Christoph  
APPLICANT: Stinson, Jeffrey R.  
APPLICANT: Uknes, Scott J.  
APPLICANT: Ward, Eric R.  
APPLICANT: Williams, Shericca C.  
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC  
TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF  
NUMBER OF SEQUENCES: 106  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CIBA-GEIGY Corporation  
STREET: 7 Skyline Drive  
CITY: Hawthorne  
STATE: New York  
COUNTRY: USA  
ZIP: 10532  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.125  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/455,416  
FILING DATE: 31-MAY-1995  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/181,271  
FILING DATE: 13-JAN-94  
APPLICATION NUMBER: US 08/093,301  
FILING DATE: 16-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/937,197  
FILING DATE: 6-NOV-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/678,378  
FILING DATE: 1-APR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/305,566  
FILING DATE: 6-FEB-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/165,667  
FILING DATE: 8-MAR-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/042,847  
FILING DATE: 6-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/632,441  
FILING DATE: 21-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/425,504  
FILING DATE: 20-OCT-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/848,506  
FILING DATE: 6-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/768,122  
FILING DATE: 27-SEP-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/580,431  
FILING DATE: 7-SEP-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/368,672  
FILING DATE: 20-JUN-1989

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/329,018  
FILING DATE: 24-MAR-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/045,957  
FILING DATE: 12-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Elmer, James Scott  
REGISTRATION NUMBER: 36,129  
REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919)541-8614  
TELEFAX: (919)541-8689  
INFORMATION FOR SEQ ID NO: 106:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1031 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-455-416-106  
Query Match 11.2% Score 30.2; DB 1; Length 1031;  
Best Local Similarity 45.3%; Pred. No. 5.5;  
Matches 110; Conservative 0; Mismatches 133; Indels 0; Gaps 0;  
QY 21 AAAGATAATGGTATAGGACGAGCTTCGATGAAATCAATGATTTTATTGAGAAATCGGT 80  
Db 256 AAACACATTGCTCTTTGTTTGTTCACACATGACATGAGTGTGCTGAAGACGAGGGT 197  
QY 81 CGGAACAGAGGGAAGAAACAAAGCCTCCCGTGGGAGAAATTCACACGGGTAAAAA 140  
Db 196 CAGTGACCGGCGACGTAATACCCGATGACCAATTTGAGACACCTGGATTAA 137  
QY 141 GGCTTTGGTAAATGGCATTATTCGGGCTTGGCAACAAAATTTCTACTATCCAG 200  
Db 136 TACTTTTGGATTTCGCAAACTCTCTTTCACCGCAACATCTGATGTTTCTACAAGCAC 77  
QY 201 GGAACGAAAGGTTACTTTTACTTTTGGATTTCGAGAGATTGGAAGACGAGGTTAT 250  
Db 76 GCAACAGAAAAGATTCCTCCATTATTTCTGAACAGCAAGCAAGATTATGATGATGCTTAT 17  
QY 261 TAT 263  
Db 16 TAT 14  
RESULT 15  
US-08-455-244-106/c  
Sequence 106, Application US/08455244  
Patent No. 5789214  
GENERAL INFORMATION:  
APPLICANT: Ryals, John A.  
APPLICANT: Alexander, Danny C.  
APPLICANT: Beck, James J.  
APPLICANT: Duesing, John H.  
APPLICANT: Friedrich, Leslie B.  
APPLICANT: Goodman, Robert M.  
APPLICANT: Harms, Christian  
APPLICANT: Meins, Jr., Frederick  
APPLICANT: Montoya, Alice  
APPLICANT: Moyer, Mary B.  
APPLICANT: Neuhaus, Jean-Marc  
APPLICANT: Payne, George B.  
APPLICANT: Sperison, Christoph  
APPLICANT: Stinson, Jeffrey R.  
APPLICANT: Uknes, Scott J.  
APPLICANT: Ward, Eric R.  
APPLICANT: Williams, Shericca C.  
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC  
TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF  
NUMBER OF SEQUENCES: 106  
CORRESPONDENCE ADDRESS:

ADDRESSEE: CIBA-GEIGY Corporation  
STREET: 7 Skyline Drive  
CITY: Hawthorne  
STATE: New York  
COUNTRY: USA  
ZIP: 10532  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/455,244  
FILING DATE: 31-MAY-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/181,271  
FILING DATE: 13-JAN-94  
APPLICATION NUMBER: US 08/093,301  
FILING DATE: 16-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/937,197  
FILING DATE: 6-NOV-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/678,378  
FILING DATE: 1-APR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/305,566  
FILING DATE: 6-FEB-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/165,667  
FILING DATE: 8-MAR-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/042,847  
FILING DATE: 6-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/632,441  
FILING DATE: 21-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/425,504  
FILING DATE: 20-OCT-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/848,506  
FILING DATE: 6-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/768,122  
FILING DATE: 27-SEP-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/580,431  
FILING DATE: 7-SEP-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/368,672  
FILING DATE: 20-JUN-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/329,018  
FILING DATE: 24-MAR-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/045,957  
FILING DATE: 12-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Elmer, James Scott  
REGISTRATION NUMBER: 36,129  
REFERENCE/DOCKET NUMBER: S-19825/F1/CGC 1727  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919)541-8614  
TELEFAX: (919)541-8689  
INFORMATION FOR SEQ ID NO: 106:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1031 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)

US-08-455-244-106

Query Match 11.2%; Score 30.2; DB 1; Length 1031;  
Best Local Similarity 45.3%; Pred. No. 5.5; Indels 0; Gaps 0;  
Matches 110; Conservative 0; Mismatches 133; Indels 0; Gaps 0;  
QY 21 AAAGATAATGGTATAGGAACGAGCTTCGATGAATCAATGATTTTATTTGAGATCGGT 80  
Db 256 AAACACATTGCTCTTGTGTTTTCACACTGAACATCGAGTTGCTGAAGACGAAGGT 197  
QY 81 CGAACAGAGGGAAGAAAAACAGCCCTCCCGTCCGGAAGATTCGAACGGGTAAAAA 140  
Db 196 CAGTGTACCGGCGACGTAAATCACCAGATGACCCATGGAACTTGGAGACTGGATTAG 137  
QY 141 GGCCTTGGTAAATTTGGCATTATTGGGCTTGGCAACAAAAATTTGAAATTTCTACTATCCAG 200  
Db 136 TACTTTTGGAAATTTGCCAAATCTCTCCTTGACCGCAACATCTCGATTGCTTCAAGCAC 77  
QY 201 GGAAACGAGAGGGTACTTTTACTTTGGATTATCGAGAGATTCGAAGAACGAGGTATT 250  
Db 76 GCACAGAAAAAGATTCCCCATTATTCTGAACAGAGAAAGATTATGATGGCTTAT 17  
QY 261 TAT 263  
Db 16 TAT 14

Search completed: November 15, 2003, 08:09:05  
Job time : 46.8989 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 15, 2003, 00:09:34 ; Search time 152.329 Seconds  
(without alignments)  
5771.357 Million cell updates/sec

Title: US-09-928-457-77  
Perfect score: 269  
Sequence: 1 CGGAGCATAAATCGTTATT.....GCARGGGTATTATCAACCG 269

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2169961 seqs, 1634102185 residues

Total number of hits satisfying chosen parameters: 4339922

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA:

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:
- 2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:
- 6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:
- 9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq:
- 10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:
- 11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:
- 12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:
- 13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:
- 14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:
- 15: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:
- 16: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:
- 17: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query %	Score	Match	Length	DB ID	Description
1	269	100.0	269	10	US-09-928-457-77	Sequence 77, Appl
2	121.8	45.3	270	10	US-09-928-457-88	Sequence 88, Appl
3	37.2	13.8	914	12	US-10-027-632-153460	Sequence 153460,
4	37.2	13.8	914	13	US-10-027-632-153460	Sequence 153460,
5	36.8	13.7	1596	14	US-10-198-846-13583	Sequence 13583, A
6	36	13.4	2000	10	US-09-938-842A-4752	Sequence 4752, Ap
7	36	13.4	185695	14	US-10-020-141-11	Sequence 11, Appl
8	36	13.4	185695	14	US-10-017-721-1	Sequence 1, Appl
9	33	12.3	1142	13	US-10-001-876-2	Sequence 2, Appl1
10	32.6	12.1	557	12	US-10-027-632-138807	Sequence 138807,
11	32.6	12.1	557	12	US-10-027-632-138808	Sequence 138808,
12	32.6	12.1	557	13	US-10-027-632-138807	Sequence 138807,
13	32.6	12.1	557	13	US-10-027-632-138808	Sequence 138808,
14	31.8	11.8	6118	12	US-10-311-455-1004	Sequence 1004, Ap
15	31.6	11.7	604	12	US-10-027-632-72923	Sequence 72923, A
16	31.6	11.7	604	12	US-10-027-632-74019	Sequence 74019, A

C 17	31.6	11.7	604	13	US-10-027-632-72923	Sequence 72923, A
C 18	31.6	11.7	604	13	US-10-027-632-74019	Sequence 74019, A
19	31.6	11.7	3673778	12	US-10-312-841-1	Sequence 1, Appl1
20	31.4	11.7	884	14	US-10-106-698-356	Sequence 356, App
21	31.4	11.7	2200	12	US-10-311-455-2164	Sequence 2164, Ap
C 21	31.2	11.6	613	12	US-10-027-632-134871	Sequence 174871,
C 23	31.2	11.6	613	13	US-10-027-632-134871	Sequence 174871,
24	31	11.5	2106	12	US-10-252-157-375	Sequence 375, App
25	31	11.5	4348	12	US-09-814-353-22000	Sequence 22000, A
26	31	11.5	5222	12	US-10-006-285-443	Sequence 443, App
27	31	11.5	186957	14	US-10-185-770-3	Sequence 3, Appl1
28	30.8	11.4	420	12	US-09-814-353-13772	Sequence 12772, A
29	30.8	11.4	421	12	US-09-814-353-5169	Sequence 5169, Ap
30	30.8	11.4	1989	10	US-09-070-927A-440	Sequence 240, App
31	30.6	11.4	598	12	US-10-027-632-238992	Sequence 238992,
32	30.6	11.4	598	13	US-10-027-632-238992	Sequence 238992,
C 33	30.6	11.4	637	12	US-10-027-632-235142	Sequence 235142,
C 34	30.6	11.4	637	13	US-10-027-632-235142	Sequence 235142,
C 35	30.6	11.4	660	14	US-10-106-698-586	Sequence 586, App
C 36	30.6	11.4	32768	10	US-09-070-927A-17	Sequence 17, Appl
C 37	30.2	11.2	497	11	US-09-764-891-6683	Sequence 6683, Ap
C 38	30.2	11.2	497	14	US-10-091-572-544	Sequence 544, App
39	30.2	11.2	1314	12	US-10-027-632-254785	Sequence 254785,
40	30.2	11.2	1314	13	US-10-027-632-254785	Sequence 254785,
C 41	30	11.2	815	12	US-10-027-632-154727	Sequence 154727,
C 42	30	11.2	815	12	US-10-027-632-154728	Sequence 154728,
C 43	30	11.2	815	13	US-10-027-632-154727	Sequence 154727,
C 44	30	11.2	815	13	US-10-027-632-154728	Sequence 154728,
45	30	11.2	6059	12	US-10-311-455-1453	Sequence 1453, Ap

ALIGNMENTS

RESULT 1

US-09-928-457-77  
; Sequence 77, Application US/09928457  
; Patent No. US20020164603A1

GENERAL INFORMATION:

; APPLICANT:

; TITLE OF INVENTION: DNA, specific proteins and peptides

; TITLE OF INVENTION: of the Neisseria meningitidis species bacteria, method  
; for obtaining them and their biological application.

; NUMBER OF SEQUENCES: 99

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30 (OEB)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/928,457

; FILING DATE: 2001-08-14

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 09/214,759

; FILING DATE: 199-12-10

; INFORMATION FOR SEQ ID NO: 77:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 269 base pairs

; TYPE: nucleotide

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

US-09-928-457-77

Query Match 100.0%; Score 269; DB 10; Length 269;

Best Local Similarity 100.0%; Pred. No. 5.5e-68;

Matches 269; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGGAGCATAAATCGTTATTAAAGATAATCGTATAGGAACGAGCTTCGATGAATCAATG 60

Db 1 CGGAGCATAAATCGTTATTAAAGATAATCGTATAGGAACGAGCTTCGATGAATCAATG 60

```

QY 61 ATTTTATTTGAGATCGTTCGGAACAGAGGGAAGAAAACAGCCTCCCGTGGCGAA 120
Db 61 ATTTTATTTGAGATCGTTCGGAACAGAGGGAAGAAAACAGCCTCCCGTGGCGAA 120
QY 121 GAATTCACACGGTAAACAAAGCCTTGTTAAATTTGGCATTATTCGGGCTTGGCAACAAA 180
Db 121 GAATTCACACGGTAAACAAAGCCTTGTTAAATTTGGCATTATTCGGGCTTGGCAACAAA 180
QY 191 TTGAATTTCTACTATCAGGGAACGAAAGGTTACTTTTACTTTTGGATTATGCAGAGA 240
Db 191 TTGAATTTCTACTATCAGGGAACGAAAGGTTACTTTTACTTTTGGATTATGCAGAGA 240
QY 241 TTCGAAGAAGCAAGGTTATTTATCAACCG 269
Db 241 TTCGAAGAAGCAAGGTTATTTATCAACCG 269

```

## RESULT 2

```

US-09-928-457-88/c
; Sequence 88, Application US/09928457
; Patent No. US20020164603A1
; GENERAL INFORMATION:

```

```

; APPLICANT:
; TITLE OF INVENTION: DNA, specific proteins and peptides
; TITLE OF INVENTION: of the Neisseria meningitidis species bacteria, method
; TITLE OF INVENTION: for obtaining them and their biological application.
; NUMBER OF SEQUENCES: 99

```

```

; COMPUTER READABLE FORM:

```

```

; MEDIUM TYPE: Floppy disk

```

```

; COMPUTER: IBM PC compatible

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS

```

```

; SOFTWARE: Patent in Release #1.0, Version #1.30 (OEB)

```

```

; CURRENT APPLICATION DATA:

```

```

; APPLICATION NUMBER: US/09/928.457

```

```

; FILING DATE: 2001-08-14

```

```

; PRIORITY APPLICATION DATA:

```

```

; APPLICATION NUMBER: 09/214,759

```

```

; FILING DATE: 199-12-10

```

```

; INFORMATION FOR SEQ ID NO: 88:

```

```

; SEQUENCE CHARACTERISTICS:

```

```

; LENGTH: 270 base pairs

```

```

; TYPE: nucleotide

```

```

; STRANDEDNESS: single

```

```

; TOPOLOGY: linear

```

```

; MOLECULE TYPE: DNA (genomic)

```

```

; HYPOTHETICAL: NO

```

```

; ANTI-SENSE: NO

```

```

US-09-928-457-88

```

```

Query Match 45.3%; Score 121.8; DB 10; Length 270;
Best Local Similarity 98.4%; Pred. No. 2.3e-25;
Matches 123; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 1 CGGACGATAAAATCGTTTAAAGATAATGTTATAGGAACGAGTTCGATGAAATCAATG 60
Db 125 CGGACGATAAAATCGTTTAAAGATAATGTTATAGGAATGAACTTCGATGAAATCAATG 66
QY 61 ATTTTATTTGAGATCGTTCGGAACAGAGGGAAGAAAACAGCCTCCCGTGGCGAA 120
Db 65 ATTTTATTTGAGATCGTTCGGAACAGAGGGAAGAAAACAGCCTCCCGTGGCGAA 6
QY 121 GAATT 125
Db 5 GAATT 1

```

## RESULT 3

```

US-10-027-632-153460
; Sequence 153460, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:

```

```

; APPLICANT: Wang, David G.

```

```

; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129

```

```

; CURRENT APPLICATION NUMBER: US/10/027,632

```

```

; CURRENT FILING DATE: 2002-04-30

```

```

; PRIOR APPLICATION NUMBER: US 60/218,006

```

```

; PRIOR FILING DATE: 2000-07-12

```

```

; PRIOR APPLICATION NUMBER: US 60/198,676

```

```

; PRIOR FILING DATE: 2000-04-20

```

```

; PRIOR APPLICATION NUMBER: US 60/193,483

```

```

; PRIOR FILING DATE: 2000-03-29

```

```

; PRIOR APPLICATION NUMBER: US 60/185,218

```

```

; PRIOR FILING DATE: 2000-02-24

```

```

; PRIOR APPLICATION NUMBER: US 60/167,363

```

```

; PRIOR FILING DATE: 1999-11-23

```

```

; PRIOR APPLICATION NUMBER: US 60/156,358

```

```

; PRIOR FILING DATE: 1999-09-28

```

```

; PRIOR APPLICATION NUMBER: US 60/146,002

```

```

; PRIOR FILING DATE: 1999-08-09

```

```

; NUMBER OF SEQ ID NOS: 325720

```

```

; SOFTWARE: FastSeq for Windows Version 4.0

```

```

; SEQ ID NO 153460

```

```

; LENGTH: 914

```

```

; TYPE: DNA

```

```

; ORGANISM: Human

```

```

US-10-027-632-153460

```

```

Query Match 13.8%; Score 37.2; DB 12; Length 914;
Best Local Similarity 50.6%; Pred. No. 1.1;
Matches 87; Conservative 1; Mismatches 84; Indels 0; Gaps 0;

```

```

QY 92 GGAGGAAAACAGCGCTCCCGTGGGAGAGATTCACCGGTAAAAAGCCTGGTAA 151
Db 137 GCAGGAAAACATGATGATAGGCAAGAGATTTGGAATGATTAATTAAGTTAC 196
QY 152 ATTGCATTATTCGGGCTTGGCAACAAAAATTTGAAATTTCTACTATCCAGGAAACGAAAG 211
Db 197 TAATAATATATTAGTAATTTCAACAAAAAGAGAAATTTAAGCTATAAGTTCAAAAACATAG 256
QY 212 GGTACTTTTACTTTTGGATTATTCAGAGATTCGAGAGACCAAGCAGGATTTAT 263
Db 257 GGAAGATTTTGGGAGAACTACCAAAATTTTCCACGAAAAACAAATATTTGT 308

```

## RESULT 4

```

US-10-027-632-153460

```

```

; Sequence 153460, Application US/10027632

```

```

; GENERAL INFORMATION:

```

```

; APPLICANT: Wang, David G.

```

```

; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129

```

```

; CURRENT APPLICATION NUMBER: US/10/027,632

```

```

; CURRENT FILING DATE: 2002-04-30

```

```

; PRIOR APPLICATION NUMBER: US 60/218,006

```

```

; PRIOR FILING DATE: 2000-07-12

```

```

; PRIOR APPLICATION NUMBER: US 60/198,676

```

```

; PRIOR FILING DATE: 2000-04-20

```

```

; PRIOR APPLICATION NUMBER: US 60/193,483

```

```

; PRIOR FILING DATE: 2000-03-29

```

```

; PRIOR APPLICATION NUMBER: US 60/185,218

```

```

; PRIOR FILING DATE: 2000-02-24

```

```

; PRIOR APPLICATION NUMBER: US 60/167,363

```

```

; PRIOR FILING DATE: 1999-11-23

```

```

; PRIOR APPLICATION NUMBER: US 60/156,358

```

```

; PRIOR FILING DATE: 1999-09-28

```

```

; PRIOR APPLICATION NUMBER: US 60/146,002

```

```

; PRIOR FILING DATE: 1999-08-09

```

```

; NUMBER OF SEQ ID NOS: 325720

```

```

; SOFTWARE: FastSeq for Windows Version 4.0

```

```

; SEQ ID NO 153460

```

```

; LENGTH: 914

```

```

; TYPE: DNA

```

```
; ORGANISM: Human
US-10-027-632-153460

Query Match      13.8%; Score 37.2; DB 13; Length 914;
Best Local Similarity 50.6%; Pred. No. 1.1;
Matches 87; Conservative 1; Mismatches 84; Indels 0; Gaps 0;

QY 92 GGAAGAAAACAGCCTCCCGTCGGAGAAATCCACGGGTAAAGAAAGCCCTGGTAA 151
DB 137 GCAAGAAAACATGAAGTAGGCCAAGAAATTTGGAAATGCAATTAATTAAGTAGTTAC 196
QY 152 ATTGGCAATTTATCGGGCTTGGCAACAAATTTGAAATTTCTACTATCCAGGGAACGAAAG 211
DB 197 TAATAATATATTAGTAATTRCAACAAACAAAGAGATTTAAGCTATATAAGTTCCAAACATAG 256
QY 212 GGTACTTTTACTTTGGATTATGAGAGATTCGAGAGCAAGGATTTAT 263
DB 257 GGAAGATTTTAGGAGAACTACCAAAATTTTCCAGGAAACAAATATTTGT 308

RESULT 5
US-10-198-846-13583
; Sequence 13583, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Yongxiao
; APPLICANT: Lillie, James
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13583
; LENGTH: 1596
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1484, 1497,
; LOCATION: 1554, 1578, 1585, 1591, 1596
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-13583

Query Match      13.7%; Score 36.8; DB 14; Length 1596;
Best Local Similarity 48.6%; Pred. No. 1.9;
Matches 101; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

QY 53 AATCAATCATTTTATTTGAGATCGGTGCGAAGGGAAGGGAAGGGAAGGCTCCCC 112
DB 1138 AAACCTCGTTTGTAGTAAAGAGGCTCGGAAGAGGAGGAGCCACAAATCTGTGTC 1197
QY 113 GTGCGGAAGAATTCACCGGTAAAGGCTTGGTAAATTTGGCAATTTTCGGGCTGG 172
DB 1198 TTCTCACAATAGTCATTTGGGCAATAAGGCATTTCTGTCTTTTGGGCTGCTGCCITCA 1257
QY 173 CAACAAATTTGAAATTTCTATCTCCAGGGAACGAAGGTTACTTTTACTTTGGATTA 232
DB 1256 GCACAGAGAGGCCCAAGAACTCTATTTCGGGGGCAACGAGGGTTAACCATCTCTTCAGTTGA 1317
QY 233 TGCAGAGATTCGAAGAAGCAAGGGTATT 260
DB 1318 ACCAGAGTTGACAAGGCCTATGGGAAT 1345

RESULT 6
US-09-938-842A-4782/c
```

```
; Sequence 4782, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRI1300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 4782
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-4782

Query Match      13.4%; Score 36; DB 10; Length 2000;
Best Local Similarity 55.6%; Pred. No. 3.5;
Matches 69; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 118 GAAGATTTCCACGGGTAAAGAAAGCCCTGGTAAATTTGCATATTCGGGGCTTGGCAACA 177
DB 767 GAACAAATAAAGAGTAAAGATTTTCATGGTAAATAGGAAATCTATTTTATTTAAAA 708
QY 178 AAATTCAAAATTTCTACTATCCAGGGAACGAAAGGTTACTTTTACTTTGGATTATGCAG 237
DB 707 AAATTCAAATACATATCTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 648
QY 238 AGAT 241
DB 647 AGTT 644

RESULT 7
US-10-020-141-11
; Sequence 11, Application US/10020141
; Publication No. US20030092013A1
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Jeanette
; APPLICANT: Ableson, Allen
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF VASCULAR DISEASE
; FILE REFERENCE: NMI-002
; CURRENT APPLICATION NUMBER: US/10/020,141
; CURRENT FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/313,097
; PRIOR FILING DATE: 2001-08-16
; PRIOR APPLICATION NUMBER: US 60/327,485
; PRIOR FILING DATE: 2001-10-05
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 185695
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-020-141-11

Query Match      13.4%; Score 36; DB 14; Length 185695;
Best Local Similarity 50.6%; Pred. No. 22;
Matches 87; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 92 GGAAGAAAACAGCCTCCCGTCGGAGAAATTCACACGGGTAAAGAAAGGCTTGGTAA 151
DB 178160 GCAAGAAAACATGAAGTAGGCCAAGAAAGATTTGGAATTCATTAATACTAGTTAC 178219
```





PRIOR APPLICATION NUMBER: US 60/218,006  
 PRIOR FILING DATE: 2000-07-12  
 PRIOR APPLICATION NUMBER: US 60/198,676  
 PRIOR FILING DATE: 2000-04-20  
 PRIOR APPLICATION NUMBER: US 60/193,483  
 PRIOR FILING DATE: 2000-03-29  
 PRIOR APPLICATION NUMBER: US 60/185,218  
 PRIOR FILING DATE: 2000-02-24  
 PRIOR APPLICATION NUMBER: US 60/167,363  
 PRIOR FILING DATE: 1999-11-23  
 PRIOR APPLICATION NUMBER: US 60/156,358  
 PRIOR FILING DATE: 1999-09-28  
 PRIOR APPLICATION NUMBER: US 60/146,002  
 PRIOR FILING DATE: 1999-08-09  
 NUMBER OF SEQ ID NOS: 325720  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 138808  
 LENGTH: 557  
 TYPE: DNA  
 ORGANISM: Human  
 US-10-027-632-138808

Query Match 12.1%; Score 32.6; DB 12; Length 557;  
 Best Local Similarity 58.9%; Pred. No. 20;  
 Matches 56; Conservative 0; Mismatches 39; Indels 0; Gaps 0;  
 QY 132 GGTAAAAAGCGCTTGGTAAATTCGCATTATTCGGGCTTGGCAACAAATTCGAAATTTCT 191  
 DB 200 GGAATAATGCGCTGGTGTTCCTCATTTGTTAAATAGCTACCAAAATTTTGTGTT 141  
 QY 192 ACTATCCAGGAAACGAAAGGTTACTTTTACTTT 226  
 DB 140 TTTCACAATTAATGCAAGATTTCATTACTCT 106

RESULT 12  
 US-10-027-632-138807/c  
 Sequence 138807, Application US/10027632  
 GENERAL INFORMATION:

APPLICANT: Wang, David G.  
 TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
 TITLE OF INVENTION: Polymorphisms in the Human Genome  
 FILE REFERENCE: 108827.129

CURRENT APPLICATION NUMBER: US/10/027,632  
 CURRENT FILING DATE: 2002-04-30  
 PRIOR APPLICATION NUMBER: US 60/218,006  
 PRIOR FILING DATE: 2000-07-12  
 PRIOR APPLICATION NUMBER: US 60/198,676  
 PRIOR FILING DATE: 2000-04-20  
 PRIOR APPLICATION NUMBER: US 60/193,483  
 PRIOR FILING DATE: 2000-03-29  
 PRIOR APPLICATION NUMBER: US 60/185,218  
 PRIOR FILING DATE: 2000-02-24  
 PRIOR APPLICATION NUMBER: US 60/167,363  
 PRIOR FILING DATE: 1999-11-23  
 PRIOR APPLICATION NUMBER: US 60/156,358  
 PRIOR FILING DATE: 1999-09-28  
 PRIOR APPLICATION NUMBER: US 60/146,002  
 PRIOR FILING DATE: 1999-08-09  
 NUMBER OF SEQ ID NOS: 325720  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 138807  
 LENGTH: 557  
 TYPE: DNA  
 ORGANISM: Human  
 US-10-027-632-138807

Query Match 12.1%; Score 32.6; DB 13; Length 557;  
 Best Local Similarity 58.9%; Pred. No. 20;  
 Matches 56; Conservative 0; Mismatches 39; Indels 0; Gaps 0;  
 QY 132 GGTAAAAAGCGCTTGGTAAATTCGCATTATTCGGGCTTGGCAACAAATTCGAAATTTCT 191

Db 200 GGAATAATGCGCTGGTGTTCCTCATTTGTTAAATAGCTACCAAAATTTTGTGTT 141  
 QY 192 ACTATCCAGGAAACGAAAGGTTACTTTTACTTT 226  
 Db 140 TTTCACAATTAATGCAAGATTTCATTACTCT 106

RESULT 13  
 US-10-027-632-138808/c  
 Sequence 138808, Application US/10027632  
 GENERAL INFORMATION:

APPLICANT: Wang, David G.  
 TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
 TITLE OF INVENTION: Polymorphisms in the Human Genome  
 FILE REFERENCE: 108827.129

CURRENT APPLICATION NUMBER: US/10/027,632  
 CURRENT FILING DATE: 2002-04-30  
 PRIOR APPLICATION NUMBER: US 60/218,006  
 PRIOR FILING DATE: 2000-07-12  
 PRIOR APPLICATION NUMBER: US 60/198,676  
 PRIOR FILING DATE: 2000-04-20  
 PRIOR APPLICATION NUMBER: US 60/193,483  
 PRIOR FILING DATE: 2000-03-29  
 PRIOR APPLICATION NUMBER: US 60/185,218  
 PRIOR FILING DATE: 2000-02-24  
 PRIOR APPLICATION NUMBER: US 60/167,363  
 PRIOR FILING DATE: 1999-11-23  
 PRIOR APPLICATION NUMBER: US 60/156,358  
 PRIOR FILING DATE: 1999-09-28  
 PRIOR APPLICATION NUMBER: US 60/146,002  
 PRIOR FILING DATE: 1999-08-09  
 NUMBER OF SEQ ID NOS: 325720  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 138808  
 LENGTH: 557  
 TYPE: DNA  
 ORGANISM: Human  
 US-10-027-632-138808

Query Match 12.1%; Score 32.6; DB 13; Length 557;  
 Best Local Similarity 58.9%; Pred. No. 20;  
 Matches 56; Conservative 0; Mismatches 39; Indels 0; Gaps 0;  
 QY 132 GGTAAAAAGCGCTTGGTAAATTCGCATTATTCGGGCTTGGCAACAAATTCGAAATTTCT 191  
 Db 200 GGAATAATGCGCTGGTGTTCCTCATTTGTTAAATAGCTACCAAAATTTTGTGTT 141  
 QY 192 ACTATCCAGGAAACGAAAGGTTACTTTTACTTT 226  
 Db 140 TTTCACAATTAATGCAAGATTTCATTACTCT 106

RESULT 14  
 US-10-311-455-1004/c  
 Sequence 1004, Application US/10311455  
 Publication No. US20030143606A1  
 GENERAL INFORMATION:

APPLICANT: OLEK, Alexander  
 APPLICANT: PIEPENBROCK, Christian  
 APPLICANT: BERLIN, Kurt  
 TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining Cytosine Methylation  
 FILE REFERENCE: 5013.1014  
 CURRENT APPLICATION NUMBER: US/10/311,455  
 CURRENT FILING DATE: 2002-12-16  
 PRIOR APPLICATION NUMBER: PCT/EP01/07537  
 PRIOR FILING DATE: 2001-07-02  
 PRIOR APPLICATION NUMBER: DE 10032529.7  
 PRIOR FILING DATE: 2000-06-30  
 PRIOR APPLICATION NUMBER: DE 10043826.1  
 PRIOR FILING DATE: 2000-09-01  
 NUMBER OF SEQ ID NOS: 2424  
 SEQ ID NO 1004

Search completed: November 15, 2003, 08:31:58  
Job time : 156.429 secs

```
; LENGTH: 6118
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1004

Query Match          11.8%; Score 31.8; DB 12; Length 6118;
Best Local Similarity 53.7%; Pred. No. 91;
Matches 66; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY 7 ATAAAAATCGTTATTAAAGATATAGTATAGGAACGAGCTTCGATGAATCAATCATTTT 66
DB 2177 AAAAAAATATAATAACAAAAAATATACGAAACTCGTCAATCTAATAATAACTTTA 2118

QY 67 ATTTGAGAAATCGTGGGACAGAGGGAAGAAAAACAAAGCTCCCGTGGGGAAGATTC 126
DB 2117 ATTAATAAATTCAAAAATATAAACCAATCATATAAAACCTTACCAATCAAAAAAATCA 2058

QY 127 CAA 129
DB 2057 TAA 2055
```

```
RESULT 15
US-10-027-632-72923/c
; Sequence 72923, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027.632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 72923
; LENGTH: 604
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-72923
```

```
Query Match          11.7%; Score 31.6; DB 12; Length 604;
Best Local Similarity 53.2%; Pred. No. 40;
Matches 67; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 115 GCGGAAGAATTCACAGCGTAAAAAGCCCTTGGTAATTCGCATTTATCGGGCTTGCA 174
DB 532 GAGGCAGAAGATCAATGCCVACTAGCTTAGGAAATATGCGACATCTCGAGAATAT 473

QY 175 ACAAAATTCGAATTTCTACTCCAGGGAACGAAAGGTTACTTTTACTTTGGATTATG 234
DB 472 TFAAAATTTAAGGAATACTAAACAAATATGAACTTGATTTCTTTTGGAACTCA 413

QY 235 CAGAGA 240
DB 412 CTGAAA 407
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 14, 2003, 23:56:24 ; Search time 1013.71 Seconds  
(without alignments)  
9241.609 Million cell updates/sec

Title: US-09-928-457-79  
Perfect score: 229  
Sequence: 1 CGGTTTCAGGTTTCGCCAA.....TGGATGCAGCATCCCAATCCG 229

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenBank:

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vl.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_on.\*
- 21: em\_ox.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vl.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_mus.\*
- 34: em\_htg\_pln.\*
- 35: em\_htg\_rod.\*
- 36: em\_htg\_mam.\*
- 37: em\_htg\_vrt.\*
- 38: em\_sy.\*
- 39: em\_hngo\_hum.\*
- 40: em\_hngo\_mus.\*
- 41: em\_hngo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	229	100.0	229	1	AF169453	AF169453 Neisseria
2	229	100.0	229	6	A68908	A68908 Sequence 79
3	229	100.0	229	6	BD063008	BD063008 DNA and s
C 4	229	100.0	349061	1	NMA22491	AL162753 Neisseria
5	227.4	99.3	15053	1	AE002550	AE002550 Neisseria
6	227.4	99.3	172325	6	AX044035	AX044035 Sequence
7	227.4	99.3	349980	6	AX044034	AX044034 Sequence
C 8	202.4	88.4	349	1	AF169440	AF169440 Neisseria
C 9	192.8	84.2	400	1	AF169418	AF169418 Neisseria
C 10	162.4	70.9	192	6	A68996	A68996 Sequence 67
C 11	162.4	70.9	192	6	BD062996	BD062996 DNA and s
12	161.2	70.4	409	1	NMAJ5415	AJ005415 Neisseria
13	107	46.7	1275	6	A96274	A96274 Sequence 30
14	107	46.7	1572	6	AX044062	AX044062 Sequence
15	107	46.7	1575	6	A96272	A96272 Sequence 30
16	107	46.7	1576	6	A96270	A96270 Sequence 30
17	100.6	43.9	1575	6	A96276	A96276 Sequence 30
18	100.6	43.9	1575	6	AX024113	AX024113 Sequence
C 19	36.8	16.1	175638	2	AL590312	AL590312 Homo sapi
20	36.6	16.0	43345	9	AL590542	AL590542 Human DNA
21	35.2	15.4	156481	2	AC079838	AC079838 Homo sapi
22	35.2	15.4	157650	9	AC007098	AC007098 Homo sapi
C 23	35.2	15.4	167711	2	AC007629	AC007629 Homo sapi
24	35	15.3	13463	2	AC013907	AC013907 Drosophil
C 25	35	15.3	164969	3	AC023693	AC023693 Drosophil
C 26	35	15.3	179312	3	AC023745	AC023745 Drosophil
C 27	35	15.3	220855	10	AC125581	AC125581 Mus muscu
C 28	35	15.3	315790	3	AE003484	AE003484 Drosophil
C 29	35	15.3	334028	2	AC116537	AC116537 Drosophil
C 30	34.6	15.1	152572	9	AC090588	AC090588 Homo sapi
C 31	34.6	15.1	152881	2	AC025737	AC025737 Homo sapi
C 32	34.6	15.1	183169	9	AC009643	AC009643 Homo sapi
C 33	34.2	14.9	137705	5	AL929326	AL929326 Zebrafish
C 34	34.2	14.9	160920	8	AC137072	AC137072 Genomic s
C 35	34.2	14.9	203558	10	AL670464	AL670464 Mouse DNA
C 36	34	14.8	175524	2	AC140955	AC140955 Didelphis
C 37	33.8	14.8	99292	9	AL445216	AL445216 Human DNA
C 38	33.6	14.7	76013	2	AP005922	AP005922 Oryza sat
C 39	33.6	14.7	99200	2	AC136895	AC136895 Leishmani
C 40	33.6	14.7	112020	2	AL133285	AL133285 Homo sapi
41	33.6	14.7	113041	9	AL391297	AL391297 Human DNA
42	33.6	14.7	138485	9	AC105273	AC105273 Homo sapi
43	33.6	14.7	141791	2	AP005582	AP005582 Oryza sat
44	33.6	14.7	206030	9	AC021439	AC021439 Homo sapi
45	33.4	14.6	1622	9	BC050472	BC050472 Homo sapi

ALIGNMENTS

RESULT 1  
AF169453  
LOCUS AF169453 229 bp DNA linear BCT 09-AUG-2000  
DEFINITION Neisseria meningitidis strain 22491 clone Cm043 unknown sequence.  
ACCESSION AF169453  
VERSION AF169453.1 GI:9754661  
KEYWORDS  
SOURCE Neisseria meningitidis  
ORGANISM Neisseria meningitidis  
Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
Neisseriaceae; Neisseria.  
REFERENCE 1 (bases 1 to 229)  
AUTHORS Perrin,A., Nassif,X. and Tinsley,C.R.  
TITLE Identification of regions of the chromosome of Neisseria meningitidis and Neisseria gonorrhoeae which are specific to

```

pathogenic Neisseriae
Unpublished
REFERENCE 2 (bases 1 to 229)
AUTHORS Perrin A., Nassif, X. and Tinsley, C.R.
TITLE Direct Submission
JOURNAL Submitted (16-JUN-1999) Necker Medicine Faculty, INSERM U411, 156
rue de Vaugirard, Paris 75015, France
FEATURES
    source
        1..229
            /organism="Neisseria meningitidis"
            /mol_type="genomic DNA"
            /strain="22491"
            /db_xref="taxon:487"
            /clone="Cm043"
BASE COUNT 72 a 48 c 65 g 44 t
ORIGIN
    Query Match 100.0%; Score 229; DB 1; Length 229;
    Best Local Similarity 100.0%; Pred. No. 5.6e-63;
    Matches 229; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGTTTCAGGTTGTCGCGAAGGCTCGGTACGGCAACCTGATTACGGGTGATCGAGCA 60
Db 1 CGTTTCAGGTTGTCGCGAAGGCTCGGTACGGCAACCTGATTACGGGTGATCGAGCA 60

Qy 61 GCTTCAACATTCGCGACCGCAAGCGGAATATGTTATCCGCAATGAGTGGCTAAAAAC 120
Db 61 GCTTCAACATTCGCGACCGCAAGCGGAATATGTTATCCGCAATGAGTGGCTAAAAAC 120

Qy 121 CAATAAAGACAATTTAGATGATGTCGGGAAGATGCGCCGACCAAGACTATGCAAAA 180
Db 121 CAATAAAGACAATTTAGATGATGTCGGGAAGATGCGCCGACCAAGACTATGCAAAA 180

Qy 181 TATGAAAAACCAAGTACCGGGATCAGGCATGCGATGCGATCCCAATCCG 229
Db 181 TATGAAAAACCAAGTACCGGGATCAGGCATGCGATGCGATCCCAATCCG 229

RESULT 2
LOCUS A68908 229 bp DNA linear PAT 06-MAY-1999
DEFINITION Sequence 79 from Patent WO9802547.
ACCESSION A68908
VERSION A68908.1 GI:4759827
KEYWORDS
SOURCE unidentified
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 229)
AUTHORS Nassif, X., Tinsley, C., Achtman, M., Ruelle, J., Vinals, C. and
Merker, P.
TITLE DNA AND SPECIFIC PROTEINS OR PEPTIDES OF THE NEISSERIA MENINGITIDIS
JOURNAL APPLICATIONS
Patent: JP 2001504684-A 70 10-APR-2001;
INSTITUT NATIONAL DE LA SANTE ET DE LA RECHERCHE MEDICALE, MAX
PLANCK GESELLSCHAFT ZUR FORDERUNG DER WISSENSCHAFTEN EV BERLIN,
SMITHKLINE BEECHAM
PN JP 2001504684-A/70
PD 10-APR-2001
PF 11-JUL-1997 JP 1998505685
PR 12-JUL-1996 FR 96/08768
PI XAVIER NASSIF, COLIN TINSLEY, MARK ACHTMAN, JEAN LOUIS RUELLE, PI
CARLA VINALS,
PI PETRA MERKER
PC C12N15/31, C07K14/22, C07K16/12, A61K39/095, C12Q1/68, G01N33/53 CC
Strandedness: Single;
CC Topology: Linear;
PH Key Location/Qualifiers.
FEATURES
    source
        1..229
            /organism="unidentified"
            /mol_type="genomic DNA"
            /db_xref="taxon:32644"
BASE COUNT 72 a 48 c 65 g 44 t
ORIGIN
    Query Match 100.0%; Score 229; DB 6; Length 229;
    Best Local Similarity 100.0%; Pred. No. 5.6e-63;
    Matches 229; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGTTTCAGGTTGTCGCGAAGGCTCGGTAAACGGCAACCTGATTACGGGTGATCGAGCA 60
Db 1 CGTTTCAGGTTGTCGCGAAGGCTCGGTAAACGGCAACCTGATTACGGGTGATCGAGCA 60

Qy 61 GTTTGAACATTCGCGACCGCAAGCGGAATATGTTTATCCGCAATGAGTGGCTAAAAAC 120
Db 61 GTTTGAACATTCGCGACCGCAAGCGGAATATGTTTATCCGCAATGAGTGGCTAAAAAC 120

Qy 121 CAATAAAGACAATTTAGATGATGTCGGGAAGATGCGCCGACCAAGACTATGCAAAA 180
Db 121 CAATAAAGACAATTTAGATGATGTCGGGAAGATGCGCCGACCAAGACTATGCAAAA 180

Qy 181 TATGAAAAACCAAGTACCGGGATCAGGCATGCGATGCGATCCCAATCCG 229
Db 181 TATGAAAAACCAAGTACCGGGATCAGGCATGCGATGCGATCCCAATCCG 229

```

```

pathogenic Neisseriae
Unpublished
REFERENCE 2 (bases 1 to 229)
AUTHORS Perrin A., Nassif, X. and Tinsley, C.R.
TITLE Direct Submission
JOURNAL Submitted (16-JUN-1999) Necker Medicine Faculty, INSERM U411, 156
rue de Vaugirard, Paris 75015, France
FEATURES
    source
        1..229
            /organism="Neisseria meningitidis"
            /mol_type="genomic DNA"
            /strain="22491"
            /db_xref="taxon:487"
            /clone="Cm043"
BASE COUNT 72 a 48 c 65 g 44 t
ORIGIN
    Query Match 100.0%; Score 229; DB 1; Length 229;
    Best Local Similarity 100.0%; Pred. No. 5.6e-63;
    Matches 229; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGTTTCAGGTTGTCGCGAAGGCTCGGTACGGCAACCTGATTACGGGTGATCGAGCA 60
Db 1 CGTTTCAGGTTGTCGCGAAGGCTCGGTACGGCAACCTGATTACGGGTGATCGAGCA 60

Qy 61 GCTTCAACATTCGCGACCGCAAGCGGAATATGTTATCCGCAATGAGTGGCTAAAAAC 120
Db 61 GCTTCAACATTCGCGACCGCAAGCGGAATATGTTATCCGCAATGAGTGGCTAAAAAC 120

Qy 121 CAATAAAGACAATTTAGATGATGTCGGGAAGATGCGCCGACCAAGACTATGCAAAA 180
Db 121 CAATAAAGACAATTTAGATGATGTCGGGAAGATGCGCCGACCAAGACTATGCAAAA 180

Qy 181 TATGAAAAACCAAGTACCGGGATCAGGCATGCGATGCGATCCCAATCCG 229
Db 181 TATGAAAAACCAAGTACCGGGATCAGGCATGCGATGCGATCCCAATCCG 229

RESULT 3
LOCUS BD063008 229 bp DNA linear PAT 27-AUG-2002
DEFINITION DNA and specific proteins or peptides of the Neisseria meningitidis
species bacteria, method for obtaining them and their biological
applications.
ACCESSION BD063008.1 GI:22608611
VERSION JP 2001504684-A/70.
KEYWORDS
SOURCE unidentified
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 229)
AUTHORS Nassif, X., Tinsley, C., Achtman, M., Ruelle, J., Vinals, C. and
Merker, P.
TITLE DNA and specific proteins or peptides of the Neisseria meningitidis
JOURNAL applications
Patent: JP 2001504684-A 70 10-APR-2001;
INSTITUT NATIONAL DE LA SANTE ET DE LA RECHERCHE MEDICALE, MAX
PLANCK GESELLSCHAFT ZUR FORDERUNG DER WISSENSCHAFTEN EV BERLIN,
SMITHKLINE BEECHAM
PN JP 2001504684-A/70
PD 10-APR-2001
PF 11-JUL-1997 JP 1998505685
PR 12-JUL-1996 FR 96/08768
PI XAVIER NASSIF, COLIN TINSLEY, MARK ACHTMAN, JEAN LOUIS RUELLE, PI
CARLA VINALS,
PI PETRA MERKER
PC C12N15/31, C07K14/22, C07K16/12, A61K39/095, C12Q1/68, G01N33/53 CC
Strandedness: Single;
CC Topology: Linear;
PH Key Location/Qualifiers.
FEATURES
    source
        1..229
            /organism="unidentified"
            /mol_type="genomic DNA"
            /db_xref="taxon:32644"
BASE COUNT 72 a 48 c 65 g 44 t
ORIGIN
    Query Match 100.0%; Score 229; DB 6; Length 229;
    Best Local Similarity 100.0%; Pred. No. 5.6e-63;
    Matches 229; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGTTTCAGGTTGTCGCGAAGGCTCGGTAAACGGCAACCTGATTACGGGTGATCGAGCA 60
Db 1 CGTTTCAGGTTGTCGCGAAGGCTCGGTAAACGGCAACCTGATTACGGGTGATCGAGCA 60

Qy 61 GTTTGAACATTCGCGACCGCAAGCGGAATATGTTTATCCGCAATGAGTGGCTAAAAAC 120
Db 61 GTTTGAACATTCGCGACCGCAAGCGGAATATGTTTATCCGCAATGAGTGGCTAAAAAC 120

Qy 121 CAATAAAGACAATTTAGATGATGTCGGGAAGATGCGCCGACCAAGACTATGCAAAA 180
Db 121 CAATAAAGACAATTTAGATGATGTCGGGAAGATGCGCCGACCAAGACTATGCAAAA 180

Qy 181 TATGAAAAACCAAGTACCGGGATCAGGCATGCGATGCGATCCCAATCCG 229
Db 181 TATGAAAAACCAAGTACCGGGATCAGGCATGCGATGCGATCCCAATCCG 229

```

```

RESULT 4
NMA222491/c
LOCUS
DEFINITION
  NMA222491 349061 bp DNA linear BCT 02-SEP-2002
  Neisseria meningitidis serogroup A strain Z2491 complete genome;
  segment 2/7.
ACCESSION
  AL162753 AL157959
VERSION
  AL162753.2 GI:7379120
KEYWORDS
  .
SOURCE
  Neisseria meningitidis Z2491
  Neisseria meningitidis Z2491
  Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
  Neisseriaceae; Neisseria.
  1 (bases 1 to 349061)
REFERENCE
  Parkhill,J., Achtman,M., James,K.D., Bentley,S.D., Churcher,C.,
  Klee,S.R., Morelli,G., Basham,D., Brown,D., Chillingworth,T.,
  Davies,R.M., Davis,P., Devlin,K., Feltwell,T., Hamlin,N.,
  Holtroyd,S., Jagels,K., Leather,S., Moule,S., Mungall,K.,
  Quail,M.A., Rajandream,M.A., Rutherford,K.M., Simmonds,M.,
  Skelton,J., Whitehead,S., Spratt,B.G. and Barrall,B.G.
  Complete DNA sequence of a serogroup A strain of Neisseria
  meningitidis Z2491
  Nature 404 (6777), 502-506 (2000)
JOURNAL
  Nature 404 (6777), 502-506 (2000)
MEDLINE
  20222556
PUBMED
  10761919
REFERENCE
  2 (bases 1 to 349061)
AUTHORS
  Parkhill,J
DIRECT SUBMISSION
  Submitted (30-MAR-2000) Submitted on behalf of the Neisseria
  sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
  Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk
  Notes:
  Details of N. meningitidis sequencing at the Sanger Centre are
  available on the World Wide Web.
  (URL, http://www.sanger.ac.uk/Projects/N_meningitidis/).
FEATURES
  source
    1..349061
    /organism="Neisseria meningitidis Z2491"
    /mol_type="genomic DNA"
    /strain="Z2491"
    /db_xref="taxon:122587"
    /note="serogroup: A"
    complement(24..206)
    /note="ATR repeat; hmms hit to HMM ATR (1 - 183), score:
    310.39"
    /label=ATR
    209..212
    /gene="NMA0368"
    220..681
    /gene="NMA0368"
    220..681
    /gene="NMA0368"
    /note="NMA0368, probable integral membrane protein, len:
    153 aa; contains four probable transmembrane domains"
    /codon_start=1
    /transl_table=11
    /product="putative integral membrane protein"
    /protein_id="CAB83669.1"
    /db_xref="GI:7379121"
    /db_xref="SPTREMBL:Q9JWH7"
    /translation="MQEONRKSPPVIMLVSLVSLVIAISLVNVAFLNNGHSGMEGLTV
    ILGSIASLDIRCAVYANYVYLAIVLALRKVKVYPVHAFWGLALVAESVKAIVV
    DEAGNTSDIVRYGAGFYLYAAFAVASIGTFAGKNKKAASADGKNDV"
    671..1492
    /gene="hemK"
    671..1492
    /gene="hemK"
    /note="NMA0369, hemK, HemK protein, len: 273 aa; similar
    to e.g. HEMK_ECOLI P37186 HEMK protein (277 aa), fasta
    scores; E(): 0, 42.3% identity in 279 aa overlap. Contains
    PS00092 N-6 Adenine-specific DNA methylases signature"
    /codon_start=1
    /transl_table=11
  repeat_unit
    /product="HemK protein"
    /protein_id="CAB83670.1"
    /db_xref="GI:7379122"
    /db_xref="SPTREMBL:Q9JWH6"
    /translation="MTFDFKWLGLSKLPKNEARMMLLOVSEYTRVQLTRGGEPMDEV
    RQADRLAORLNGEVAIVLIGAREFYGRFTVNPVSLIPRPETELHVAVLARLPEN
    GRWDVLTGSGAVAVTVAFRPAFVRSADISPPALETKNAADLCARVAFYGSWF
    DTMDPSGKWDIIIVSNPFIENGKHLSDGLRFEFCIALTDFSDGLSCIRTLAQGAP
    DRLAEGGFLLEHGFQGAARVGLAENGFSGVETLPDLAAGLDRVTLGKYMHLK"
    831..840
    /gene="hemK"
    /note="Core DNA uptake sequence: gccgtctgaa"
    /label=DUS
    1162..1171
    /gene="hemK"
    /note="Core DNA uptake sequence: gccgtctgaa"
    /label=DUS
    1187..1207
    /gene="hemK"
    /note="PS00092 N-6 Adenine-specific DNA methylases
    signature"
    complement(1279..1288)
    /note="Core DNA uptake sequence: gccgtctgaa"
    /label=DUS
    1565..1568
    /gene="NMA0370"
    1578..2966
    /gene="NMA0370"
    1578..2966
    /gene="NMA0370"
    /note="NMA0370, probable integral membrane protein, len:
    462 aa; similar to hypothetical proteins e.g. Y325_HAEIN
    P44640 hypothetical protein HI0325 (450 aa), fasta_scores;
    E(): 0, 49.8% identity in 464 aa overlap"
    /codon_start=1
    /transl_table=11
    /product="putative integral membrane protein"
    /protein_id="CAB83671.1"
    /db_xref="GI:7379123"
    /db_xref="SPTREMBL:Q9JWH5"
    /translation="MNAVVAIVMLVLSLRVHVVLSTIGAFVGGAVAGMPLONIA
    DAQOVQAGIIPVFNKLEGGAIALSVAMLGAFAMATHSGLPQOLAGAVVRKLR
    GMPDSVRSGEVTKWLLSILVWGMSSNIPIHIAFIPIVFPFLLVNNAIKLDR
    RLIACVITGLVTYMFPGFAGIFNEILNIGHSIIHSAAPQLDQKXINVNAALPAIL
    MYAGLLAFHYRKPRLYQSNNAADNAGNARFPQSAVRSALAAVAIAVCFALQ
    MVDLSVLGAMLGFAFMVLGVIINRDKANDVFGEGIKMMAMVGFIMIAAQGFVAVMNA
    TGHQIQLVSSMAIFGNSKGMALMVVGLLVTMGIGSFSFTPIIAIYVPLCVGL
    GSPLATVAIVGTAGALGDAGSPASDSTIGPTMGLNADGQHDHRSVPTFIHNP
    LLIAAGNIAAWL"
    2967..3191
    /gene="slyX"
    2967..3191
    /gene="slyX"
    /gene="slyX"
    /note="NMA0371, slyX, SLIX protein homolog, len: 74 aa;
    similar to SLIX_HAEIN P44759 SLIX protein homolog (73 aa),
    fasta_scores; E(): 0.0017, 33.8% identity in 74 aa,
    overlap, and SLIX_ECOLI P30857 SLIX protein (72 aa), fasta
    scores; E(): 0.24, 32.4% identity in 68 aa overlap"
    /codon_start=1
    /transl_table=11
    /product="SLIX protein homolog"
    /protein_id="CAB83672.1"
    /db_xref="GI:7379124"
    /db_xref="SPTREMBL:Q9JWH4"
    /translation="MDAVQEFHRIETELEIQSALQEDVIAGLNAMVAELRQTLDLQQA
    QLRLYXQMDRNPDAQEYSLRDEIPHY"
    complement(3271..3423)
    /gene="NMA0372"
    complement(3271..3423)
    /gene="NMA0372"
    /note="NMA0372, unknown, questionable CDS, len: 50 aa"
    /codon_start=1
    /transl_table=11
    /product="very hypothetical protein NMA0372"

```

```

misc_feature
/protein_id="CAB83673.1"
/db_xref="GI:7379125"
/db_xref="SPTREMBL:Q5JWH3"
/translation="WFGRLSPGSDGFFLAVYEPVGVICLQGTAYFRSGGKRFHRI
KGIISD"
3562...3571
/note="Core DNA uptake sequence: gcgcgtctgaa"
/label=DUS
/gene="thiP"
/complement(3586..4356)
/gene="thiP"
/complement(3586..4356)
/gene="thiP"
/note="NMA0373, thiP, probable ThiF protein, len: 256 aa;
similar to e.g. THIF_ECOLI P30138 ThiF protein (251 aa),
fasta scores; E(): 0, 43.1% identity in 246 aa overlap,
and MOEB_ECOLI P12282 molybdopterin biosynthesis MOEB
protein_ (249 aa), fasta scores; E(): 0, 43.9% identity in
244 aa overlap (note that N.m. does not have orthologs of
any other molybdopterin biosynthesis proteins). Contains
Pfam match to entry PF00899 ThiF_family, ThiF family"
/codon_start=1
/transl_table=11
/product="ThiF protein"
/protein_id="CAB83674.1"
/db_xref="GI:7379126"
/db_xref="SPTREMBL:Q5JWH2"
/translation="MTTEHNDDAFLRYSRHLLDEIGEGQOKLSAAHILVVGCG
HGLAALPYLAASGILTIADSDTVELHNLQVAFDEGVDGKLTALADLRHIN
HTVDRTNKLDCRITGLQVAAIDVLDCDNATQAVNACVQAKTPLVSGAAVR
FEGGLAVRPLDPSGVACILFDGSGASDGICSLFGVFSPLVIGISTQAAEALKILL
DAGPSHGLAVRALEGWGYFDLPNPECPVCAER"
3787...3796
/note="Core DNA uptake sequence: gcgcgtctgaa"
/label=DUS
/gene="thiP"
/complement(3847..4257)
/note="Pfam match to entry PF00899 ThiF_family, ThiF
family, score 186.60, E-value 4.1e-52"
/complement(4418..4427)
/note="Core DNA uptake sequence: gcgcgtctgaa"
/label=DUS
/gene="ppc"
/EC numbers="4.1.1.31"
/note="NMA0374, ppc, phosphoenolpyruvate carboxylase, len:
917 aa; similar to many e.g. CAPP_RHOA Q32483
phosphoenolpyruvate carboxylase (EC 4.1.1.31) (936 aa),
fasta scores; E(): 0, 43.3% identity in 928 aa overlap.
Contains 2x Pfam match to entry PF00311 PEPCase,
Phosphoenolpyruvate carboxylase, PS00017 ATP/GTP-binding
site motif A (P-loop), and PS00393 Phosphoenolpyruvate
carboxylase active site 2"
/codon_start=1
/transl_table=11
/product="phosphoenolpyruvate carboxylase"
/protein_id="CAB83675.1"
Query Match 100.0%; Score 229; DB 1; Length 349061;
Best Local Similarity 100.0%; Pred. No. 1.7e-62;
Matches: 229; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGTTTTCAGTTGTCGGAAGCTCGGTACGGCAACCTGATTACGGGTGATCAGGCA 60
Db 64614 CGTTTTCAGTTGTCGGAAGCTCGGTACGGCAACCTGATTACGGGTGATCAGGCA 64555
Qy 61 GCTTGAACATTCGCAAGCGGAGGGAATATGTTATCCCATGAGTCGCGTAAAAAC 120
Db 64554 GCTTGAACATTCGCAAGCGGAGGGAATATGTTATCCCATGAGTCGCGTAAAAAC 64495
Qy 121 CAATAAAGACAAATTAGATGATGTCGGGAAGATGCCCGACCGCAAGACTATGCAAAA 180

```

```

Db 64494 CAATAAAGACAAATTAGATGATGTCGGGAAGATGCCCGACCGCAAGACTATGCAAAA 64435
Qy 181 TATGAAAAACCAAGTACGCGGATCAGGCATGAGCATGATGATCAATCCG 229
Db 64434 TATGAAAAACCAAGTACGCGGATCAGGCATGAGCATGATGATCAATCCG 64386

RESULT 5
AE002550 15053 bp DNA linear BCT 25-MAY-2000
LOCUS Neisseria meningitidis serogroup B strain MC58 section 192 of 206
DEFINITION of the complete genome.
ACCESSION AE002550 A5002098
VERSION AE002550.2 GI:7413476
KEYWORDS Neisseria meningitidis MC58
SOURCE Neisseria meningitidis MC58
ORGANISM Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
Neisseriaceae; Neisseria.
REFERENCE 1 (bases 1 to 15053)
AUTHORS Tettelin,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C.,
Nettleson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.F.,
Dodson,R.J., Nelson,W.C., Winn,M.L., DeBoy,R., Peterson,J.D.,
Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D.,
Dougherty,B.A., Mason,T., Ciecko,A., Parksey,D.S., Blair,E.,
Cittone,H., Clark,E.B., Cotton,M.D., Utterback,T.R., Khouri,H.,
Qin,H., Vamathevan,J., Gill,J., Scarlato,V., Masignani,V.,
Pizza,M., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R.,
Rappuoli,R. and Venter,J.C.
Complete genome sequence of Neisseria meningitidis serogroup B
strain MC58
Science 287 (5459), 1809-1815 (2000)
JOURNAL 20175755
MEDLINE 10710307
PUBMED
REFERENCE 2 (bases 1 to 15053)
AUTHORS Tettelin,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C.,
Nettleson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.F.,
Dodson,R.J., Nelson,W.C., Winn,M.L., DeBoy,R., Peterson,J.D.,
Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D.,
Dougherty,B.A., Mason,T., Ciecko,A., Parksey,D.S., Blair,E.,
Cittone,H., Clark,E.B., Cotton,M.D., Utterback,T.R., Khouri,H.,
Qin,H., Vamathevan,J., Gill,J., Scarlato,V., Masignani,V.,
Pizza,M., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R.,
Rappuoli,R. and Venter,J.C.
Direct Submission
TITLE Submitted (17-MAR-2000) The Institute for Genomic Research, 9712
JOURNAL Medical Center Dr. Rockville, MD 20850, USA
COMMENT On Apr 4, 2000 this sequence version replaced gi:7227263.
FEATURES
source
1..15053
/organism="Neisseria meningitidis MC58"
/mol_type="genomic DNA"
/strain="MC58"
/db_xref="taxon:122586"
/note="serogroup: B"
/complement(74..982)
/gene="NMB2000"
/complement(74..982)
/gene="NMB2000"
/note="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF42327.1"
/db_xref="GI:7227264"
/translation="MNQTAINRADVTRTFIDMEVRLHVRLENWQHIVKQKNYPA
ARRALGELLAAVLLSGNLKNEGTLIVQVQGRGLKMLVAEASDRVTRATARWDET
ASIADESGLDLGEGGVFLTLPKDGEPKQGVVPLEGGIAQMLVNMKSEBOLDT
HVLVSASAEAGGLVQLPEVLDSEAEHNVSTILRTLREELAGLDAQHVLFLFH
ETPPVFPFETFEFSCITSRGKVSMDLLMLDGEVGVVVEGSGSIEVDCDFCHSKYIF
DETVDNALFGEDEVGVAKGLPRHTVQ"
1227..1982
gene

```

CDS  
 /gene="NMB2001"  
 1227..11982  
 /note="NMB2001"  
 /notes="conserved hypothetical protein; identified by  
 Glimmer2; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="AAF42328.1"  
 /db\_xref="GI:7227265"  
 /translation="MDSFKPAVAVLWLVFAVRLADLTLLSSREQILRPAED  
 EQVLPIPARARAGNADDELIGSAMGLEQVLPVNPVARRAGNABELIGNANGLN  
 EQVLPIPARARAGNADDELIGSAMGLEQVLPVNPVARRAGNABELIGNANGLN  
 GNLPTAEQRMGTIPVARSLEQPMQDMVFRTLGSRISHVGLYIGNRRFIHAPRTG  
 KNIETISLHKYWSGYAPAREVKKNDPSRLN"  
 2001..2219  
 /gene="NMB2002"  
 2001..2219  
 /gene="NMB2002"  
 /note="hypothetical protein; identified by Glimmer2;  
 putative"  
 /transl\_table=11  
 /codon\_start=1  
 /product="hypothetical protein"  
 /protein\_id="AAF42329.1"  
 /db\_xref="GI:7227266"  
 /translation="MSMEMPKAVYDDQIVSCTEKKVVMSENMAELYQTQAQAFEDA  
 LMGGERQLRDYLLALIEGLENPYRVK"  
 2364..2711  
 /gene="NMB2003"  
 2364..2711  
 /gene="NMB2003"  
 /note="conserved hypothetical protein; identified by  
 Glimmer2; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="AAF42330.1"  
 /db\_xref="GI:7227267"  
 /translation="WNIRALLIILGLGTGETAVFLAGIKLPQSIQVGMVLPALLOA  
 GWKTSWQQITDALMSNLTLFVPPCVAVISYLDLIADDFWSILVSASASTLCVLIV  
 IGKVRHWIRGIIR"  
 2708..3400  
 /gene="NMB2004"  
 2708..3400  
 /gene="NMB2004"  
 /note="conserved hypothetical protein; identified by  
 Glimmer2; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="AAF42331.1"  
 /db\_xref="GI:7227268"  
 /translation="MNSLLRQPSVLLFLTLAVYALAIIVTRTGNIFCNPLVSTIVL  
 VTKLIGIDFVAVHNAQDFDFWLKPAVVVLAPELVNRRKIFNQLPVIQSLAGS  
 VTGVTGYNLGAEREVLSLASKSVTNPIAIEITRSIGGIPAITAATVIAIGIV  
 QIAGYKMLKNTVMPSSVSGSLGTASHANGIAASLERSRRAAYAGLITFNGVLTA  
 LIAPLLIPVLGF"  
 3467..4687  
 /gene="NMB2005"  
 3467..4687  
 /gene="NMB2005"  
 /note="similar to GB:M65216 SP:P38434 GB:S85363 PID:150243  
 PID:246444 percent identity: 98.77; identified by sequence  
 similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="glutamate N-acetyltransferase/amino-acid  
 acetyltransferase"  
 /protein\_id="AAF42332.1"  
 /db\_xref="GI:7227269"  
 /translation="WAVNLAKTEAEQLPDIDGIALYTAQAGVKKPGHDTLTLAVAG  
 STVGAVFTTNRFPCAPVHIAKSHLFDEGVRALVINTGNANAGTGAQGRIDALAVCAA

AARQIGCKPNOVLFFSTGVILEPLPADKIIAALPKMQPAPWNEAARAAMTTDTVPKAA  
 SREGKVGDKHTVATGIAKSGMHPNMAIPLGIATDAKVSQVPLQMTQEIADTFE  
 NTITVDGDTSTNDSFVIAATGKNSQSEIDNIADPRTAQLKELCSLAELAQAIVRDG  
 EGATKFTIVRENARKTDEARQAAYASPLVKTAFASDPNGLRGAALAAIGVAGVAD  
 LDDLVEMYLDDILVAEHGGRASAYTEAQGVAMSKAEITVRIKLHRGQAATVYTCD  
 LSHGYVSINADYRS"  
 4747..6156  
 /gene="NMB2006"  
 4747..6156  
 /gene="NMB2006"  
 /notes="similar to PID:1652876 percent identity: 52.04;  
 identified by sequence similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="chloride channel protein-related protein"  
 /protein\_id="AAF42333.1"  
 /db\_xref="GI:7227270"  
 /translation="MSTFPRLARKIRQTRRSKSIAPFLFLLAGSALVALTALFFA  
 HLADEFALNKLIVQYVFAWVPLGLFPLIAWTRKFAFPFAGSGIPQVIAISLIP  
 YGAQKTRILRLQTLIKIPLTFLGMLFGASIGREGPSVQVGAAMGAWCKHGLIA  
 FKGQENDLMAAGAGLAAAFNPLAGVIFAIEELGRGIMLRWQIILGLVASQFI  
 QVAIOGNPYPSFGVGLLEHIFLWVALSGVCGAAGLGRLLYVGAFAFAFRKIRG  
 FIRNPILLAAALMGLLALLGTFTVQKTYGTGYHEAAQALHGIYEAPGLAAGKWLAT  
 VFSYWAGVPGGIPTPSLITGAVLGEHIAIAIDISQGANIILICMAAFLAGAQCSEIT  
 SAVVMEITGGQSLFLFWMLTACIFASQVSRQSPRPFFHASGMRFRQVLOETAAQTG  
 NAPPAPQTANSKTMPSEN"  
 6346..7737  
 /gene="NMB2007"  
 /notes="This region contains a match to at least one other  
 gene that is not full length, and is not the result of a  
 sequencing artifact; similar to GP:1742299 percent  
 identity: 79.36; identified by sequence similarity;  
 putative"  
 7881..9524  
 /gene="NMB2008"  
 7881..9524  
 /gene="NMB2008"  
 /notes="similar to PID:862629 percent identity: 53.03;  
 identified by sequence similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="ABC transporter, ATP-binding protein-related  
 protein"  
 /protein\_id="AAF42334.1"  
 /db\_xref="GI:7227271"  
 /translation="MDEIQIPKKVELQTKLENEKIVLSKGSTIIIVANGTKRLAV  
 YIEQLKEXAHRISAHRAKLPNVNKIPEKAKTIVSYQNWGDIGVNRKNYRWDN  
 NSYTHLLNDFWLLQVLFQAQNNIAVANNKLNKNEKVTNKTLDILOEAWETLLPH  
 RKLHITADDIOVSADVNBELYSASNMDSGERALFVILGOVLSVDGSLVDFEPELPH  
 HXSLISNLWDKIEELPDCSFLIITHDIEPATRVAKKXVIRNYIPTPAWDISEVDES  
 NDEBITIWLRSRPIFLVEGNNSLDIATRYCIPDWIIIPKGRCKDVIQSVSELK  
 KUSNPLNLKCSGIVDUDSREIEQLNNLGIYLPVSEIENLFSUTDVAKEILK  
 LNQYSEELNKLNGFKSELIKYIDNELKDDLFVVKQVRKIDNYLKNIDLSKI  
 TETDMKSLINEISTLEQKIEIETWSEIKNEIQRCIEQQDLKLLIYDNKGLLAKSA  
 CVLKGMRNKHFEFSWIMRTLKGRNKFDAIRQKPLID"  
 9568..9750  
 /gene="NMB2009"  
 /notes="This region contains a gene with one or more  
 premature stops or frameshifts, and is not the result of a  
 sequencing artifact; similar to SP:Q09530 PID:733602  
 percent identity: 61.70; identified by sequence  
 similarity; putative"  
 9962..11539

Query Match 99.3%; Score 227.4; DB 1; Length 15053;  
 Best Local Similarity 99.6%; Pred. No. 3.4e-62;  
 Matches 228; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 CGGTTTCAGGTTGTGCGAAGGCTGGTAACCGGCAACCTGATTACGGGTGATGAGGCA 60  
 |||||  
 DB 11433 CGGTTTCAGGTTGTGCGAAGGCTGGTAACCGGCAACCTGATTACGGGTGATGAGGCA 11492  
 |||||  
 QY 61 GCTTGAACATTCGCGACGCAAGGCGGAATATGTTTATCCGCAATGATGGCGTAAAAAC 120





```

/strain="Z2491"
/db_xref="taxon:487"
/clone="Em013"
BASE COUNT      76 a      97 c      85 g      91 t
ORIGIN

Query Match      88.4%; Score 202.4; DB 1; Length 349;
Best Local Similarity 99.5%; Pred. No. 2.4e-54;
Matches 203; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGTTTCAGGTTGTCGGAAGGCTCGGTAAACGGGCAACCTGATACGGGTGATGACGGCA 60
D 1 |||||
D 204 CGTTTCAGGTTGTCGGAAGGCTCGGTAAACGGGCAACCTGATACGGGTGATGACGGCA 145
QY 61 GCTTGAACATTCGCGACGCGGAGGCGGAATATGTTTATCCGCAATGAGTGGCTAAAC 120
D 144 GCTTGAACATTCGCGACGCGGAGGCGGAATATGTTTATCCGCAATGAGTGGCTAAAC 85
QY 121 CAATAAGACAAATTTAGATGATGTCGGGGAAGATGCCCGACCGCAAGAGATATGCAAAA 180
D 84 CAATAAGACAAATTTAGATGATGTCGGGGAAGAGCGCCGACCGCAAGAGATATGCAAAA 25
QY 181 TATGAAACCAAGTACCGGATC 204
D 24 TATGAAACCAAGTACCGGATC 1

RESULT 9
AF169418/c
LOCUS
DEFINITION Neisseria gonorrhoeae strain FAL090 clone Bg007 unknown sequence.
ACCESSION AF169418
VERSION AF169418.1 GI:9754626
KEYWORDS
SOURCE
ORGANISM
Neisseria gonorrhoeae
Neisseria gonorrhoeae
Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
Neisseriaceae; Neisseria.
REFERENCE
1 (bases 1 to 400)
Perrin,A., Nassif,X. and Tinsley,C.R.
Identification of regions of the chromosome of Neisseria
meningitidis and Neisseria gonorrhoeae which are specific to
pathogenic Neisseriae
Unpublished
2 (bases 1 to 400)
Perrin,A., Nassif,X. and Tinsley,C.R.
Direct Submission
Submitted (16-JUL-1999) Necker Medicine Faculty, INSERM U411, 156
rue de Vaugirard, Paris 75015, France
LOCATION/Qualifiers
1..400
/organism="Neisseria gonorrhoeae"
/mol_type="genomic DNA"
/strain="FAL090"
/db_xref="taxon:485"
/clone="Bg007"
BASE COUNT      87 a      104 c      98 g      111 t
ORIGIN

Query Match      84.2%; Score 192.8; DB 1; Length 400;
Best Local Similarity 96.6%; Pred. No. 3.2e-51;
Matches 197; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 CGTTTCAGGTTGTCGGAAGGCTCGGTAAACGGGCAACCTGATACGGGTGATGACGGCA 60
D 204 CGTTTCAGGTTGTCGGAAGGCTCGGTAAACGGGCAACCTGATACGGGTGATGACGGCA 145
QY 61 GCTTGAACATTCGCGACGCGGAGGCGGAATATGTTTATCCGCAATGAGTGGCTAAAC 120
D 144 GCTTGAACATTCGCGACGCGGAGGCGGAATATGTTTATCCGCAATGAGTGGCTAAAC 85
QY 121 CAATAAGACAAATTTAGATGATGTCGGGGAAGATGCCCGACCGCAAGAGATATGCAAAA 180
D 181 TATGAAACCAAGTACCGGATC 204
D 24 TATGAAACCAAGTACCGGATC 1

/strain="Z2491"
/db_xref="taxon:487"
/clone="Em013"
BASE COUNT      76 a      97 c      85 g      91 t
ORIGIN

Query Match      88.4%; Score 202.4; DB 1; Length 349;
Best Local Similarity 99.5%; Pred. No. 2.4e-54;
Matches 203; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGTTTCAGGTTGTCGGAAGGCTCGGTAAACGGGCAACCTGATACGGGTGATGACGGCA 60
D 1 |||||
D 204 CGTTTCAGGTTGTCGGAAGGCTCGGTAAACGGGCAACCTGATACGGGTGATGACGGCA 145
QY 61 GCTTGAACATTCGCGACGCGGAGGCGGAATATGTTTATCCGCAATGAGTGGCTAAAC 120
D 144 GCTTGAACATTCGCGACGCGGAGGCGGAATATGTTTATCCGCAATGAGTGGCTAAAC 85
QY 121 CAATAAGACAAATTTAGATGATGTCGGGGAAGATGCCCGACCGCAAGAGATATGCAAAA 180
D 84 CAATAAGACAAATTTAGATGATGTCGGGGAAGAGCGCCGACCGCAAGAGATATGCAAAA 25
QY 181 TATGAAACCAAGTACCGGATC 204
D 24 TATGAAACCAAGTACCGGATC 1

RESULT 9
AF169418/c
LOCUS
DEFINITION Neisseria gonorrhoeae strain FAL090 clone Bg007 unknown sequence.
ACCESSION AF169418
VERSION AF169418.1 GI:9754626
KEYWORDS
SOURCE
ORGANISM
Neisseria gonorrhoeae
Neisseria gonorrhoeae
Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
Neisseriaceae; Neisseria.
REFERENCE
1 (bases 1 to 400)
Perrin,A., Nassif,X. and Tinsley,C.R.
Identification of regions of the chromosome of Neisseria
meningitidis and Neisseria gonorrhoeae which are specific to
pathogenic Neisseriae
Unpublished
2 (bases 1 to 400)
Perrin,A., Nassif,X. and Tinsley,C.R.
Direct Submission
Submitted (16-JUL-1999) Necker Medicine Faculty, INSERM U411, 156
rue de Vaugirard, Paris 75015, France
LOCATION/Qualifiers
1..400
/organism="Neisseria gonorrhoeae"
/mol_type="genomic DNA"
/strain="FAL090"
/db_xref="taxon:485"
/clone="Bg007"
BASE COUNT      87 a      104 c      98 g      111 t
ORIGIN

Query Match      84.2%; Score 192.8; DB 1; Length 400;
Best Local Similarity 96.6%; Pred. No. 3.2e-51;
Matches 197; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 CGTTTCAGGTTGTCGGAAGGCTCGGTAAACGGGCAACCTGATACGGGTGATGACGGCA 60
D 204 CGTTTCAGGTTGTCGGAAGGCTCGGTAAACGGGCAACCTGATACGGGTGATGACGGCA 145
QY 61 GCTTGAACATTCGCGACGCGGAGGCGGAATATGTTTATCCGCAATGAGTGGCTAAAC 120
D 144 GCTTGAACATTCGCGACGCGGAGGCGGAATATGTTTATCCGCAATGAGTGGCTAAAC 85
QY 121 CAATAAGACAAATTTAGATGATGTCGGGGAAGATGCCCGACCGCAAGAGATATGCAAAA 180
D 181 TATGAAACCAAGTACCGGATC 204
D 24 TATGAAACCAAGTACCGGATC 1

RESULT 10
A68896/c
LOCUS
DEFINITION Sequence 67 from Patent WO9802547.
ACCESSION A68896
VERSION A68896.1 GI:4759815
KEYWORDS
SOURCE
ORGANISM
unidentified
unidentified
unclassified.
REFERENCE
1 (bases 1 to 192)
Nassif,X., Tinsley,C., Achtman,M., Ruelle,J., Vinals,C. and
Merker,P.
DNA AND SPECIFIC PROTEINS OR PEPTIDES OF THE NEISSERIA MENINGITIDIS
SPECIES BACTERIA, METHOD FOR OBTAINING THEM AND THEIR BIOLOGICAL
APPLICATIONS
JOURNAL
Patent: WO 9802547-A 67 22-JAN-1998;
INST NAT SANTE RECH MED (FR)
Other publication FR 2751000 19980116.
FEATURES
Location/Qualifiers
1..192
/source
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"
BASE COUNT      38 a      51 c      39 g      64 t
ORIGIN

Query Match      70.9%; Score 162.4; DB 6; Length 192;
Best Local Similarity 93.8%; Pred. No. 2e-41;
Matches 180; Conservative 0; Mismatches 11; Indels 1; Gaps 1;

QY 14 TCGCGAAGGCTCGGTAAACGGGCAACCTGATACGGGTGATGACGGGTGATGACATTGCG 73
D 192 TCTCGAAGGCTCGGTAAACGGGCAACCTGATACGGGTGATGACGGGTGATGACATTGCG 133
QY 74 CGACGGCAAGCGGGAATA-TGTTATCCGCAATGAGTGGGTAACCAATGAAACAA 132
D 132 CGACGGCAAGCGGGAATA-TGTTATCCGCAATGAGTGGGTAACCAATGAAACAA 73
QY 133 ATTATAGATGATGTCGGGGAAGATGCCCGACCGCAAGAGTATGCAAAATGAAAAACCA 192
D 72 ATTATAGATGATGTCGGGGAAGATGCCCGACCGCAAGAGTATGCAAAATGAAAAACCA 13
QY 193 AGTACGGGATC 204
D 12 AGTACGGGATC 1

RESULT 11
BD062996/c
LOCUS
DEFINITION BD062996
ACCESSION BD062996
VERSION BD062996.1 GI:22608599
KEYWORDS
SOURCE
ORGANISM
unidentified
unclassified.
REFERENCE
1 (bases 1 to 192)
Nassif,X., Tinsley,C., Achtman,M., Ruelle,J.L., Vinals,C. and
Merker,P.
DNA AND SPECIFIC PROTEINS OR PEPTIDES OF THE NEISSERIA MENINGITIDIS
SPECIES BACTERIA, METHOD FOR OBTAINING THEM AND THEIR BIOLOGICAL
APPLICATIONS.
JOURNAL
JP 2001504684-A/58.
BASE COUNT      192 bp
ORIGIN
DNA and specific proteins or peptides of the Neisseria meningitidis
species bacteria, method for obtaining them and their biological
applications.

```

JOURNAL Patent: JP 2001504684-A 58 10-APR-2001;  
 INSTITUT NATIONAL DE LA SANTE ET DE LA RECHERCHE MEDICALE, MAX  
 PLANCK GESELLSCHAFT ZUR FORDERUNG DER WISSENSCHAFTEN EV BERLIN,  
 SMITHKLINE BEECHAM.  
 COMMENT PN JP 2001504684-A/58  
 PD 10-APR-2001  
 PF 11-JUL-1997 JP 19980505685  
 PR 12-JUL-1996 FR 96/08768  
 PI XAVIER NASSIF, COLIN TINSLEY, MARK ACHTMAN, JEAN LOUIS RUELLE, PI  
 CARLA VINALS,  
 PETRA MERKER  
 PC C12N15/31, C07K14/22, C07K16/12, A61K39/095, C12Q1/68, G01N33/53 CC  
 Strandedness: Single;  
 CC Topology: Linear;  
 FH Key Location/Qualifiers.  
 FEATURES source  
 1..192  
 /organism="unidentified"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:32644" 64 t  
 BASE COUNT 38 a 51 c 39 g 64 t  
 ORIGIN  
 Query Match 70.9%; Score 162.4; DB 6; Length 192;  
 Best Local Similarity 93.8%; Pred. No. 2e-41;  
 Matches 180; Conservative 0; Mismatches 11; Indels 1; Gaps 1;  
 QY 14 TCGCAAGCTCGGTAAACGGGCAACCTGATTACGGGTGATCGAGCGAGCTTGAACATTCG 73  
 Db 192 TCTCGAAGCTCGGTAAACGGGCAACCTGATTACGGGTGATCGAGCTAGCTAGCTGACATTCG 133  
 QY 74 CGACGGCAAGCGGGAATA-TGTTATTCGCAATGAGTGGCGTAAACCAATAAAGACAA 132  
 Db 132 CGACGGCAAGCGGGAATATGTTATTCGCAATGAGTGGCGTAAACCAATAAAGACAA 73  
 QY 133 ATTAGATGATGTCGGGAAGATCCCGACGACGACTATGCAAAATATGAAACCA 192  
 Db 72 ATTAGATGATGTCGGGAAGATCCCGACGACGACTATGCAAAATATGAAACCA 13  
 QY 193 AGTACGGGATC 204  
 Db 12 AGTACGGGATC 1  
 RESULT 12  
 NMAJ5415 409 bp DNA linear BCT 15-DEC-1998  
 LOCUS  
 DEFINITION Neisseria meningitidis partial DNA sequence, clone hrtB.  
 ACCESSION AJ005415  
 VERSION AJ005415.1 GI:3212117  
 KEYWORDS  
 SOURCE Neisseria meningitidis  
 ORGANISM Neisseria meningitidis  
 Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
 Neisseriaceae; Neisseria.  
 REFERENCE 1  
 AUTHORS Claus, H., Froesch, M. and Vogel, U.  
 TITLE Identification of a hotspot for transformation of Neisseria meningitidis by shuttle mutagenesis using signature-tagged transposons  
 JOURNAL Mol. Gen. Genet. 259 (4), 363-371 (1998)  
 MEDLINE 99005251  
 PUBMED 9730590  
 REFERENCE 2 (bases 1 to 409)  
 AUTHORS Vogel, U.  
 TITLE Direct Submission  
 JOURNAL Submitted (06-APR-1998) Vogel U., University of Wuerzburg,  
 Institute for Hygiene and Microbiology, Josef-Schneider-Str. 2,  
 97080 Wuerzburg, GERMANY  
 COMMENT Partial sequence of meningococcal locus giving rise to high transformation frequency.  
 FEATURES source  
 Location/Qualifiers  
 1..409

/organism="Neisseria meningitidis"  
 /mol\_type="genomic DNA"  
 /strain="B1940"  
 /isolate="German"  
 /db\_xref="taxon:487"  
 /clone="hrtB"  
 BASE COUNT 111 a 98 c 105 g 91 t 4 others  
 ORIGIN  
 Query Match 70.4%; Score 161.2; DB 1; Length 409;  
 Best Local Similarity 93.0%; Pred. No. 5.4e-41;  
 Matches 199; Conservative 0; Mismatches 12; Indels 3; Gaps 3;  
 QY 1 CGGTTTCAGGTTGTCGGAAGGCTCGGTAACGGGCAACCTGATTACGGGTGATCGAGCA 60  
 Db 197 CGGTTTCAGGTTGTCGGAAGGCTCGGTAACGGGCAACCTGATTACGGGTGATCGAGCA 256  
 QY 61 GCTTGAACATTCGGAACGGCAAGCGGGAATATCTTTATCCGCAATGAGTGGC-TAAAAA 119  
 Db 257 GCTTGAGCATTCGCGATGCGCAAGCGGGAATATGTTTATCCGCAATGAGTGGCTTAAAAA 316  
 QY 120 CCAATAAAGACAAATTTAGATGATGTCGGGGAAGATGCCGACCGCAAGACTATGCAAA 179  
 Db 317 CCAATAAAGACAAATTTAGATGATGTCGGGGAAGANGCCGACCGACAA-ANTATGCCAA 375  
 QY 180 ATATGAAAAACCAAGTACCGGATCAGGCATCGA 213  
 Db 376 ATATGAAAAACCAA-TTCNCGATNAGGCATCGA 408  
 RESULT 13  
 A96274 1275 bp DNA linear PAT 07-SEP-2000  
 LOCUS  
 DEFINITION Sequence 307 from Patent WO924578.  
 ACCESSION A96274  
 VERSION A96274.1 GI:6780029  
 KEYWORDS  
 SOURCE unidentified  
 ORGANISM unidentified  
 unclassified.  
 REFERENCE 1  
 AUTHORS Pizza, M., Scarlato, V., Rappuoli, R., Grandi, G. and Maignani, V.  
 TITLE Neisserial antigens  
 JOURNAL Patent: WO 924578-A 307 20-MAY-1999;  
 PIZZA VARIAGRAZIA (IT); SCARLATO VINCENZO (IT); RAPPUOLI RINO (IT);  
 CHIRON SPA (IT); GRANDI GUIDO (IT); MASNIGNANI VEGA (IT)  
 FEATURES  
 source  
 Location/Qualifiers  
 1..1275  
 /organism="unidentified"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:32644"  
 BASE COUNT 300 a 304 c 319 g 352 t  
 ORIGIN  
 Query Match 46.7%; Score 107; DB 6; Length 1275;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-23;  
 Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CGGTTTCAGGTTGTCGGAAGGCTCGGTAACGGGCAACCTGATTACGGGTGATCGAGCA 60  
 Db 1169 CGGTTTCAGGTTGTCGGAAGGCTCGGTAACGGGCAACCTGATTACGGGTGATCGAGCA 1228  
 QY 61 GCTTGAACATTCGGAACGGCAAGCGGGAATATGTTTATCCGCAATGA 107  
 Db 1229 GCTTGAACATTCGGAACGGCAAGCGGGAATATGTTTATCCGCAATGA 1275  
 RESULT 14  
 AX024062 1572 bp DNA linear PAT 15-SEP-2000  
 LOCUS  
 DEFINITION Sequence 5 from Patent FR2785293.  
 ACCESSION AX024062  
 VERSION AX024062.1 GI:10184374

KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
JOURNAL  
PASTEUR MERIEUX SERUMS VACC (FR)  
FEATURES  
source  
1..1572  
/organism="Neisseria meningitidis"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:487"  
1..1572  
/note="unnamed protein product"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="CAC08907.1"  
/db\_xref="GI:10184375"  
/translation="MKKSLFLVFLYSSLLTASEIAYRFVFGIETLPAKVAETPALTF  
MIAALYLPARYKASRLLIAPFARSIIANNVHYAVIQSWMTGINYMLKEITEVGSA  
GASMDKWLWLVLMVLFCSLAKRKHFSADILFAPLMIFVRSFDTKQEH  
GISPKPTYSRIKANYFSEGVFVRLPYQLDLRIAPKQAPSKIQGSVQNVILI  
MGESEAAHLKLFYGRFSLPFLRLSQADFKIVKOSYAGFMVSLPSFFNAIPH  
ANGLEQISGGTNNFRLAKEQGYETYFSAQANEMAILNLIGKWKIDHILQIPTOLCY  
GNGDNMPDEKLLPLFDKINLQGRHFIVLHGRGSHAPYSALLOPQDKVGEGLIVDKYD  
NTIHKTDOMITVFEOLOKQPDGNKFLAYTSDHQYVRQDIYNOGTVPQPSYLVPLVL  
YSSNKAVQQAANQAPCEIAFHQQLSTFLIHLIGYDMPVSGCREGVSVTGNLITGDAG  
SLNRDGRAEYIFQ"  
BASE COUNT 380 a 383 c 395 g 414 t  
ORIGIN  
Query Match 46.7%; Score 107; DB 6; Length 1572;  
Best Local Similarity 100.0%; Pred. No. 2.3e-23;  
Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CGGTTTCAGGTTGTCGGAAGGCTCGGTACGGGCAACCTGATTACGGGTGATCGAGGCA 60  
DB 1466 CGGTTTCAGGTTGTCGGAAGGCTCGGTACGGGCAACCTGATTACGGGTGATCGAGGCA 1525  
QY 61 GCTTGAACATTCCGACGCGCAAGCGGAATATGTTTATCCGCAATGA 107  
DB 1526 GCTTGAACATTCCGACGCGCAAGCGGAATATGTTTATCCGCAATGA 1572  
RESULT 15  
A96272  
LOCUS A96272 1575 bp DNA linear PAT 07-SEP-2000  
DEFINITION Sequence 305 from Patent WO924578.  
ACCESSION A96272  
VERSION A96272.1 GI:6780028  
KEYWORDS  
SOURCE unidentified  
ORGANISM unidentified  
REFERENCE  
AUTHORS Pizza.M., Scarlato.V., Rappuoli.R., Grandi.G. and Masignani.V.  
TITLE Neisserial antigens  
JOURNAL Patent: WO 924578-A 305 20-MAY-1999;  
PIZZA MARIAGRAZIA (IT); SCARLATO VINCENZO (IT); RAPPUOLI RINO (IT);  
CHIRON SPA (IT); GRANDI GUIDO (IT); MASIGNANI VEGA (IT)  
FEATURES  
source  
1..1575  
/organism="unidentified"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32644"  
BASE COUNT 376 a 378 c 405 g 416 t  
ORIGIN  
Query Match 46.7%; Score 107; DB 6; Length 1575;  
Best Local Similarity 100.0%; Pred. No. 2.3e-23;

Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CGGTTTCAGGTTGTCGGAAGGCTCGGTACGGGCAACCTGATTACGGGTGATCGAGGCA 60  
DB 1469 CGGTTTCAGGTTGTCGGAAGGCTCGGTACGGGCAACCTGATTACGGGTGATCGAGGCA 1528  
QY 61 GCTTGAACATTCCGACGCGCAAGCGGAATATGTTTATCCGCAATGA 107  
DB 1529 GCTTGAACATTCCGACGCGCAAGCGGAATATGTTTATCCGCAATGA 1575  
Search completed: November 15, 2003, 03:33:55  
Job time : 1016.71 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 14, 2003, 23:55:19 ; Search time 147.797 Seconds  
(without alignments)  
4182.570 Million cell updates/sec

Title: US-09-928-457-79  
Perfect score: 229  
Sequence: 1 CGGTTTCAGGTGTGCGAA.....TGGATGCAGCATCAATCCG 229

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 252756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq 19Jun03.\*

1: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*  
2: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*  
3: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*  
4: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*  
5: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*  
6: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*  
7: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*  
8: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*  
9: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*  
10: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*  
11: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*  
12: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*  
13: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*  
14: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*  
15: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*  
16: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*  
17: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*  
18: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*  
19: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*  
20: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*  
21: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*  
22: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*  
23: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*  
24: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*  
25: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.\*

RESULT	Query	Match	Length	ID	Description
1	229	100.0	229	19	AAV03587
2	227.4	99.3	12438	21	AAA81550
3	227.4	99.3	172325	21	AAF21613
4	227.4	99.3	345980	21	AAF21612
5	227.4	99.3	837096	21	AAA81489
6	182.4	70.9	192	19	AAV03576
7	107	46.7	366	21	AAA81358
8	107	46.7	624	20	AAZ12104

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	Match	Length	ID	Description
1	229	100.0	229	19	AAV03587
2	227.4	99.3	12438	21	AAA81550
3	227.4	99.3	172325	21	AAF21613
4	227.4	99.3	345980	21	AAF21612
5	227.4	99.3	837096	21	AAA81489
6	182.4	70.9	192	19	AAV03576
7	107	46.7	366	21	AAA81358
8	107	46.7	624	20	AAZ12104

RESULT	Query	Match	Length	ID	Description
1	229	100.0	229	19	AAV03587
2	227.4	99.3	12438	21	AAA81550
3	227.4	99.3	172325	21	AAF21613
4	227.4	99.3	345980	21	AAF21612
5	227.4	99.3	837096	21	AAA81489
6	182.4	70.9	192	19	AAV03576
7	107	46.7	366	21	AAA81358
8	107	46.7	624	20	AAZ12104

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	Match	Length	ID	Description
1	229	100.0	229	19	AAV03587
2	227.4	99.3	12438	21	AAA81550
3	227.4	99.3	172325	21	AAF21613
4	227.4	99.3	345980	21	AAF21612
5	227.4	99.3	837096	21	AAA81489
6	182.4	70.9	192	19	AAV03576
7	107	46.7	366	21	AAA81358
8	107	46.7	624	20	AAZ12104

Neisseria meningitidis  
DNA encoding a pol  
Neisseria meningitidis  
Neisseria meningitidis  
DNA encoding a pol  
N. gonorrhoeae nuc  
N. gonorrhoeae nuc  
N. gonorrhoeae nuc  
Human secreted exp  
Hypoxanthine phosph  
Human CDNA encodin  
Human novel polynu  
CDNA encoding novel  
DNA sequence #188  
Human coding sequ  
Drosophila melanog  
Drosophila melanog  
Yeast AOD9604-asso  
Oligonucleotide fo  
Oligonucleotide fo  
Aspergillus oryzae  
Human lung tumour-  
Human lung tumour-  
Human CDNA clone (  
Human CDNA encodin  
Human CDNA #43 dif  
Human CDNA sequenc  
Human lung tumour-  
Human lung tumour-  
Human ovarian anti  
Human polynucleoti  
DNA encoding human  
Human reproductive  
Human immune/haema  
Human immune/haema  
Aspergillus oryzae



diagnosis; antigen; detection; infection; gene therapy; antibacterial; ds.

Neisseria meningitidis.

WO200066791-A1.

09-NOV-2000.

08-MAR-2000; 2000WO-US05928.

30-APR-1999; 99US-0132068.

08-OCT-1999; 99WO-US23573.

28-FEB-2000; 2000GB-0004695.

(CHIR ) CHIRON CORP.

(GENO-) INST GENOMIC RES.

Pizza M, Hickey E, Peterson J, Tettelin H, Venter JC, Massignani V; Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V, Rappuoli R; Frazer CM, Grandi G;

WPI; 2000-647603/62.

Neisseria meningitidis B full length genome sequence and open reading frames are used to detect, treat and prevent Neisserial infections -

Claim 7; Appendix A; 692pp; English.

The present invention describes the full length genome of Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607 to AAF21613 represent fragments of the NMB genomic sequence, as the sequence was too long to go in a record on its own it was split into 8 sequences which overlap each other at the beginning and end of each sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at the beginning of AAF21607, the last 49980 bp of AAF21607 are repeated at the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the Neisseria proteins given in AAB58550 to AAB58593, and AAF21589 to AAF21606 represent PCR primers which are used in the exemplification of the present invention. The NMB genome and fragments from it have antibacterial activity, and can be used in vaccines and gene therapy. Neisseria nucleic acids, proteins and/or antibodies which binds to the proteins can be used in compositions for treating or preventing infection due to Neisserial bacteria or as a diagnostic reagent for detecting the presence of Neisserial bacteria or of antibodies raised to Neisserial bacteria. Computers, computer memory, computer storage medium or computer databases can be used in a search to identify open reading frames (ORFs) or coding sequences within the NMB genome. The DNA sequences provide further opportunities to find antigenic or immunogenic proteins which are more effective in vaccines than the outer membrane proteins currently used.

Sequence 172325 BP; 43072 A; 47583 C; 41465 G; 40205 T; 0 other;

Query Match 99.3%; Score 227.4; DB 21; Length 172325;  
Best Local Similarity 99.6%; Pred. No. 58-66;  
Matches 228; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 CGGTTTCAGGTTGCGGAGCGCTCGGTAACGGGCAACCTGATTACGGGTGATGCGGCA 60  
24545 CGGTTTCAGGTTGCGGAGCGCTCGGTAACGGGCAACCTGATTACGGGTGATGCGGCA 24604

61 GCTTGAACATTCGGACGGCGGAGCGGAATATGTTTATCGCAATGATGCGGTAAAC 120  
24605 GCTTGAACATTCGGACGGCGGAGCGGAATATGTTTATCGCAATGATGCGGTAAAC 24664

121 CAATAAAGACAAATTTAGATGATCGGGGAAGATGCCGACCGACAGACTGACAAA 180  
24665 CAATAAAGACAAATTTAGATGATCGGGGAAGATGCCGACCGACAGACTGACAAA 24724

181 TATGAAAAACCAAGTACCGGATCAGCATGGATGCGACGATCCAAATCCG 229  
24725 TATGAAAAACCAAGTACCGGATCAGCATGGATGCGACGATCCAAATCCG 24773

RESULT 4  
AAF21612  
ID AAF21612 standard; DNA; 349980 BP.  
XX  
AC AAF21612;  
XX  
DT 13-MAR-2001 (first entry)  
XX  
DE Neisseria meningitidis B nucleotide sequence SEQ ID NO:113.  
XX  
KW Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;  
KW diagnosis; antigen; detection; infection; gene therapy; antibacterial;  
KW ds.  
XX  
OS Neisseria meningitidis.  
XX  
PN WO200066791-A1.  
XX  
PD 09-NOV-2000.  
XX  
PF 08-MAR-2000; 2000WO-US05928.  
XX  
PR 30-APR-1999; 99US-0132068.  
PR 08-OCT-1999; 99WO-US23573.  
PR 28-FEB-2000; 2000GB-0004695.  
XX  
XX (CHIR ) CHIRON CORP.  
PA (GENO-) INST GENOMIC RES.  
PA  
PI Pizza M, Hickey E, Peterson J, Tettelin H, Venter JC, Massignani V;  
PI Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V, Rappuoli R;  
PI Frazer CM, Grandi G;  
XX  
XX WPI; 2000-647603/62.  
XX  
XX Neisseria meningitidis B full length genome sequence and open reading  
PT frames are used to detect, treat and prevent Neisserial infections -  
PT  
XX  
PS Claim 7; Appendix A; 692pp; English.  
XX  
XX The present invention describes the full length genome of  
CC Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607  
CC to AAF21613 represent fragments of the NMB genomic sequence, as the  
CC sequence was too long to go in a record on its own it was split into 8  
CC sequences which overlap each other at the beginning and end of each  
CC sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at  
CC the beginning of AAF21607, the last 49980 bp of AAF21607 are repeated at  
CC the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the  
CC Neisseria proteins given in AAB58550 to AAB58593, and AAF21589 to  
CC AAF21606 represent PCR primers which are used in the exemplification of  
CC the present invention. The NMB genome and fragments from it have  
CC antibacterial activity, and can be used in vaccines and gene therapy.  
CC Neisseria nucleic acids, proteins and/or antibodies which binds to the  
CC proteins can be used in compositions for treating or preventing infection  
CC due to Neisserial bacteria or as a diagnostic reagent for detecting the  
CC presence of Neisserial bacteria or of antibodies raised to Neisserial  
CC bacteria. Computers, computer memory, computer storage medium or computer  
CC databases can be used in a search to identify open reading frames (ORFs)  
CC or coding sequences within the NMB genome. The DNA sequences provide  
CC further opportunities to find antigenic or immunogenic proteins which are  
CC more effective in vaccines than the outer membrane proteins currently  
CC used.  
XX  
SQ Sequence 349980 BP; 86473 A; 95646 C; 85908 G; 81953 T; 0 other;

Query Match 99.3%; Score 227.4; DB 21; Length 349980;  
Best Local Similarity 99.6%; Pred. No. 6.6e-66;  
Matches 228; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 CGGTTTCAGGTTGCGGAGCGCTCGGTAACGGGCAACCTGATTACGGGTGATGCGGCA 60  
|||||





The present invention describes methods of obtaining immunogenic proteins from *Neisseria* genomic sequences. AA81453 to AA82414 represent specifically claimed *Neisseria meningitidis* genomic DNA sequences, AA81260 to AA81303 and AA825820 to AA825663 represent *Neisseria* DNA sequences and their corresponding proteins; AA81254 to AA81259 and AA81304 to AA81321 represent PCR primers used in the isolation of *Neisseria meningitidis* DNA sequences; and AA81322 to AA81452 represent *Neisseria meningitidis* MenB polynucleotide ORF sequences, which are all used in the exemplification of the present invention. The nucleic acid sequences, protein sequences, and antibodies against them, can be used in the manufacture of a composition. The composition can be used as a medicament for in the manufacture of a medicament for treating, preventing or diagnosing infection due to *Neisseria* bacteria. For example, some of the identified proteins could be components of vaccines against *Meningococcus B*; against all serotypes, and/or against all pathogenic *Neisseriae*. Identification of sequences from the bacterium will also facilitate production of biological probes, particularly organism-specific probes. Attempts to make efficacious *Meningococcus B* vaccines have failed mainly due to antigen tolerance. Multivalent vaccines have also been tried but none have successfully overcome antigenic variability. The provision of further, complete sequences may provide an opportunity to identify secreted or surface exposed proteins that may be presumed targets for the immune system and which are not antigenically variable or at least more conserved than other more variable regions.

Query Match 46.7%; Score 107; DB 21; Length 366;  
 Best Local Similarity 100.0%; Pred. NO. 3.4e-26;  
 Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Db      260  CGGTTTCAGGTTGTCGGGAAGGTCGGTAACGGGCAACCTGATTACGGGTGATCGAGCA 319
QY      61   GCTTGAACATTCCGACGCGCAAGCGGAATATGTTTATCGCAATGA 107
Db      320  GCTTGAACATTCCGACGCGCAAGCGGAATATGTTTATCGCAATGA 366

RESULT 8
AAZ12104
ID      AAZ12104 standard; DNA; 624 BP.
XX
AC      AAZ12104;
XX
XX      08-OCT-1999 (first entry)
XX
XX      Neisseria meningitidis partial ORF81 sequence.
XX
XX      Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
XX      treatment; Neisseria infection; meningitis; septicaemia; gonorrhea; ss.
XX
XX      Neisseria meningitidis.
XX
XX      Key      Location/Qualifiers
XX      misc_feature      256..257
XX      FT      /tag= a
XX      FT      /note= 952 unspecified nucleotides are present
XX      FT      between these bases"
XX
XX      W09924578-A2.
XX
XX      PN
XX      PD
XX      PD
XX      PD
XX      PF      09-OCT-1998; 98WO-IB01665.
XX
XX      01-SEP-1998; 98GB-0019016.
XX      06-NOV-1997; 97GB-0023516.
XX      14-NOV-1997; 97GB-0024190.
XX      18-NOV-1997; 97GB-0024386.

```

```

PR 27-NOV-1997; 97GB-0025159.
PR 10-DEC-1997; 97GB-0026147.
PR 14-JAN-1998; 98GB-0000759.
XX (CHIR-) CHIRON SPA.
PI Grandi G, Masignani V, Pizza M, Rappuoli R, Scarlato V;
XX WPI; 1999-327407/27.
DR P-PSDB; AAY38650.
XX Proteins from Neisseria meningitidis and N. Gonorrhoeae useful for
PT diagnosis, treatment and prevention of infection
XX Claim 9; Page 204; 524pp; English.
XX Nucleotide sequences AA211972-Z12358 represent open reading frames
CC (ORFs) of Neisseria meningitidis and N. gonorrhoeae which encode
CC antigenic proteins (see AAY38499-Y38944). The antigenic proteins, their
CC fragments, their nucleic acids and antibodies are used for diagnosis,
CC prevention (as vaccines) or treatment of Neisseria infections,
CC such as meningitis, septicemia and gonorrhea. Both organisms
CC are closely related. Fragments of the nucleic acids are useful
CC as hybridisation probes and antisense reagents.
XX Claim 9; Page 204; 524pp; English.
XX Nucleotide sequences AA211972-Z12358 represent open reading frames
CC (ORFs) of Neisseria meningitidis and N. gonorrhoeae which encode
CC antigenic proteins (see AAY38499-Y38944). The antigenic proteins, their
CC fragments, their nucleic acids and antibodies are used for diagnosis,
CC prevention (as vaccines) or treatment of Neisseria infections,
CC such as meningitis, septicemia and gonorrhea. Both organisms
CC are closely related. Fragments of the nucleic acids are useful
CC as hybridisation probes and antisense reagents.
XX Query Match 46.7%; Score 107; DB 20; Length 624;
Best Local Similarity 100.0%; Pred. No. 4.1e-26;
Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX 1 CGGTTTCAGGTTGTCGCGAAGCTCGGTAACGGCAACCTGATTACGGGTGATCGAGCA 60
Db 518 CGGTTTCAGGTTGTCGCGAAGCTCGGTAACGGCAACCTGATTACGGGTGATCGAGCA 577
XX 61 GCTTGAACATTCGCGACGCGAAGCTCGGTAACGGCAACCTGATTACGGGTGATCGAGCA 107
Db 578 GCTTGAACATTCGCGACGCGAAGCTCGGTAACGGCAACCTGATTACGGGTGATCGAGCA 624
RESULT 9
AA212106
ID AA212106 standard; DNA; 1275 BP.
XX AC AA212106;
XX DT 08-OCT-1999 (first entry)
XX DE Neisseria meningitidis strain A complete ORF81 sequence.
XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW treatment; Neisseria infection; meningitis; septicaemia; gonorrhea; ss.
XX OS Neisseria meningitidis.
XX PN WO9924578-A2.
XX PD 20-MAY-1999.
XX PF 09-OCT-1999; 98WO-IB01665.
XX PR 01-SEP-1998; 98GB-0019016.
XX PR 06-NOV-1997; 97GB-0023516.
XX PR 14-NOV-1997; 97GB-0024190.
XX PR 18-NOV-1997; 97GB-0024386.
XX PR 27-NOV-1997; 97GB-0025158.
XX PR 10-DEC-1997; 97GB-0026147.
XX PR 14-JAN-1998; 98GB-0000759.
XX (CHIR-) CHIRON SPA.
PI Grandi G, Masignani V, Pizza M, Rappuoli R, Scarlato V;
XX WPI; 1999-327407/27.
DR P-PSDB; AAY38650.
XX Proteins from Neisseria meningitidis and N. Gonorrhoeae useful for
PT diagnosis, treatment and prevention of infection
XX Claim 9; Page 204; 524pp; English.
XX Nucleotide sequences AA211972-Z12358 represent open reading frames
CC (ORFs) of Neisseria meningitidis and N. gonorrhoeae which encode
CC antigenic proteins (see AAY38499-Y38944). The antigenic proteins, their
CC fragments, their nucleic acids and antibodies are used for diagnosis,
CC prevention (as vaccines) or treatment of Neisseria infections,
CC such as meningitis, septicemia and gonorrhea. Both organisms
CC are closely related. Fragments of the nucleic acids are useful
CC as hybridisation probes and antisense reagents.
XX Query Match 46.7%; Score 107; DB 20; Length 624;
Best Local Similarity 100.0%; Pred. No. 4.1e-26;
Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX 1 CGGTTTCAGGTTGTCGCGAAGCTCGGTAACGGCAACCTGATTACGGGTGATCGAGCA 60
Db 518 CGGTTTCAGGTTGTCGCGAAGCTCGGTAACGGCAACCTGATTACGGGTGATCGAGCA 577
XX 61 GCTTGAACATTCGCGACGCGAAGCTCGGTAACGGCAACCTGATTACGGGTGATCGAGCA 107
Db 578 GCTTGAACATTCGCGACGCGAAGCTCGGTAACGGCAACCTGATTACGGGTGATCGAGCA 624
RESULT 9
AA212106
ID AA212106 standard; DNA; 1275 BP.
XX AC AA212106;
XX DT 08-OCT-1999 (first entry)
XX DE Neisseria meningitidis strain A complete ORF81 sequence.
XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW treatment; Neisseria infection; meningitis; septicaemia; gonorrhea; ss.
XX OS Neisseria meningitidis.
XX PN WO9924578-A2.
XX PD 20-MAY-1999.
XX PF 09-OCT-1999; 98WO-IB01665.
XX PR 01-SEP-1998; 98GB-0019016.
XX PR 06-NOV-1997; 97GB-0023516.
XX PR 14-NOV-1997; 97GB-0024190.
XX PR 18-NOV-1997; 97GB-0024386.
XX PR 27-NOV-1997; 97GB-0025158.
XX PR 10-DEC-1997; 97GB-0026147.
XX PR 14-JAN-1998; 98GB-0000759.
XX (CHIR-) CHIRON SPA.
PI Grandi G, Masignani V, Pizza M, Rappuoli R, Scarlato V;
XX WPI; 1999-327407/27.
DR P-PSDB; AAY38652.
XX Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT diagnosis, treatment and prevention of infection
XX Claim 9; Page 206; 524pp; English.
XX Nucleotide sequences AA211972-Z12358 represent open reading frames
CC (ORFs) of Neisseria meningitidis and N. gonorrhoeae which encode
CC antigenic proteins (see AAY38499-Y38944). The antigenic proteins, their
CC fragments, their nucleic acids and antibodies are used for diagnosis,
CC prevention (as vaccines) or treatment of Neisseria infections,
CC such as meningitis, septicemia and gonorrhea. Both organisms
CC are closely related. Fragments of the nucleic acids are useful
CC as hybridisation probes and antisense reagents.
XX Query Match 46.7%; Score 107; DB 20; Length 1275;
Best Local Similarity 100.0%; Pred. No. 5.5e-26;
Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX 1 CGGTTTCAGGTTGTCGCGAAGCTCGGTAACGGCAACCTGATTACGGGTGATCGAGCA 60
Db 1169 CGGTTTCAGGTTGTCGCGAAGCTCGGTAACGGCAACCTGATTACGGGTGATCGAGCA 1228
XX 61 GCTTGAACATTCGCGACGCGAAGCTCGGTAACGGCAACCTGATTACGGGTGATCGAGCA 107
Db 1229 GCTTGAACATTCGCGACGCGAAGCTCGGTAACGGCAACCTGATTACGGGTGATCGAGCA 1275
RESULT 10
AA215298
ID AA215298 standard; DNA; 1572 BP.
XX AC AA215298;
XX DT 04-SEP-2000 (first entry)
XX DE DNA encoding a polypeptide of a Neisseria pathogenic strain.
XX Pathogenic strain; Neisseria; vaccine; Neisseria infection; ss.
XX OS Neisseria meningitidis.
XX FH Key Location/Qualifiers
FT CDS 1..1572
FT FT /*tag= a
XX WO200026375-A2.
XX PN 11-MAY-2000.
XX PD 28-OCT-1999; 99WO-FR02643.
XX PF 30-OCT-1998; 98FR-0013693.
XX PR (INMR ) PASTEUR MERIEUX SERUMS & VACCINS SA.
PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX PI Aujame L, Bouchardon A, Renauld-Mongenie G, Rokbi B, Nassif X;
XX Tinsley C, Perrin A;
XX WPI; 2000-365622/31.
DR P-PSDB; AAY93267.
XX New polypeptide specific for pathogenic Neisseria useful in therapeutic
PT or preventative vaccines and for diagnosis
XX Claim 2; Page 56-59; 187pp; French.
XX The present sequence encodes a protein that is specific for pathogenic

```

CC strains of *Neisseria*. The polynucleotides, polypeptides, or their  
 CC antigenic fragments, are used in vaccines to treat or protect against  
 CC *Neisseria* infections, particularly by *N. meningitidis*. The  
 CC polynucleotide sequence is also used for recombinant production of  
 CC the polypeptide and to produce attenuated *Neisseria* strains that  
 CC overexpress it, or express it in a non-toxic mutant form.  
 XX  
 SQ Sequence 1572 BP; 380 A; 383 C; 395 G; 414 T; 0 other;

Query Match 46.7%; Score 107; DB 21; Length 1572;  
 Best Local Similarity 100.0%; Pred. No. 5.9e-26;  
 Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGTTTCAGGTTGTCGCGAAGGCTCGGTAAACGGCAACCTGATTACGGGTGATCGAGGCA 60  
 DB 1466 CGTTTCAGGTTGTCGCGAAGGCTCGGTAAACGGCAACCTGATTACGGGTGATCGAGGCA 1525

QY 61 GCTTGACATTCGCGACGGCAAGCGGGAATATGTTTATCCGCAATGA 107  
 DB 1526 GCTTGACATTCGCGACGGCAAGCGGGAATATGTTTATCCGCAATGA 1572

RESULT 11  
 AAZ12105  
 ID AAZ12105 standard; DNA; 1575 BP.  
 XX  
 AC AAZ12105;  
 XX  
 DT 08-OCT-1999 (first entry)  
 XX  
 DE *Neisseria meningitidis* complete ORF81 sequence.  
 XX  
 KW *Neisseria meningitidis*; *Neisseria gonorrhoeae*; antigen; vaccine;  
 KW treatment; *Neisseria* infection; meningitis; septicaemia; gonorrhea; ss.  
 XX  
 OS *Neisseria meningitidis*.  
 XX  
 PN WO9924578-A2.  
 XX  
 PD 20-MAY-1999.  
 XX  
 PF 09-OCT-1998; 98WB-IB01665.  
 XX  
 PR 01-SEP-1998; 98GB-0019016.  
 PR 06-NOV-1997; 97GB-0023516.  
 PR 14-NOV-1997; 97GB-0024190.  
 PR 18-NOV-1997; 97GB-0024386.  
 PR 27-NOV-1997; 97GB-0025158.  
 PR 10-DEC-1997; 97GB-0026147.  
 PR 14-JAN-1998; 98GB-0000759.  
 XX  
 PA (CHIR-) CHIRON SPA.  
 XX  
 PI Grandi G, Masignani V, Pizza M, Rappuoli R, Scarlato V;  
 XX  
 DR WPI; 1999-327407/27.  
 DR P-PSDB; AAY38651.  
 XX  
 XX Proteins from *Neisseria meningitidis* and *N. gonorrhoeae* useful for  
 PT diagnosis, treatment and prevention of infection  
 XX  
 PS Claim 9; Page 205; 524pp; English.  
 XX  
 CC Nucleotide sequences AAZ11972-212358 represent open reading frames  
 CC (ORFs) of *Neisseria meningitidis* and *N. gonorrhoeae* which encode  
 CC antigenic proteins (see AAY38499-Y38944). The antigenic proteins, their  
 CC fragments, their nucleic acids and antibodies are used for diagnosis,  
 CC prevention (as vaccines) or treatment of *Neisseria* infections,  
 CC such as meningitis, septicaemia and gonorrhea. Both organisms  
 CC are closely related. Fragments of the nucleic acids are useful  
 CC as hybridisation probes and antisense reagents.  
 XX  
 SQ Sequence 1575 BP; 376 A; 378 C; 405 G; 416 T; 0 other;

Query Match 46.7%; Score 107; DB 20; Length 1575;  
 Best Local Similarity 100.0%; Pred. No. 5.9e-26;  
 Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGTTTCAGGTTGTCGCGAAGGCTCGGTAAACGGCAACCTGATTACGGGTGATCGAGGCA 60  
 DB 1469 CGTTTCAGGTTGTCGCGAAGGCTCGGTAAACGGCAACCTGATTACGGGTGATCGAGGCA 1528

QY 61 GCTTGACATTCGCGACGGCAAGCGGGAATATGTTTATCCGCAATGA 107  
 DB 1529 GCTTGACATTCGCGACGGCAAGCGGGAATATGTTTATCCGCAATGA 1575

RESULT 12  
 AAZ12107  
 ID AAZ12107 standard; DNA; 1575 BP.  
 XX  
 AC AAZ12107;  
 XX  
 DT 08-OCT-1999 (first entry)  
 XX  
 DE *Neisseria gonorrhoeae* complete ORF81 sequence.  
 XX  
 KW *Neisseria meningitidis*; *Neisseria gonorrhoeae*; antigen; vaccine;  
 KW treatment; *Neisseria* infection; meningitis; septicaemia; gonorrhea; ss.  
 XX  
 OS *Neisseria gonorrhoeae*.  
 XX  
 PN WO9924578-A2.  
 XX  
 PD 20-MAY-1999.  
 XX  
 PF 09-OCT-1998; 98WO-IB01665.  
 XX  
 PR 01-SEP-1998; 98GB-0019016.  
 PR 06-NOV-1997; 97GB-0023516.  
 PR 14-NOV-1997; 97GB-0024190.  
 PR 18-NOV-1997; 97GB-0024386.  
 PR 27-NOV-1997; 97GB-0025158.  
 PR 10-DEC-1997; 97GB-0026147.  
 PR 14-JAN-1998; 98GB-0000759.  
 XX  
 PA (CHIR-) CHIRON SPA.  
 XX  
 PI Grandi G, Masignani V, Pizza M, Rappuoli R, Scarlato V;  
 XX  
 DR WPI; 1999-327407/27.  
 DR P-PSDB; AAY38653.  
 XX  
 XX Proteins from *Neisseria meningitidis* and *N. gonorrhoeae* useful for  
 PT diagnosis, treatment and prevention of infection  
 XX  
 PS Claim 9; Page 208; 524pp; English.  
 XX  
 CC Nucleotide sequences AAZ11972-212358 represent open reading frames  
 CC (ORFs) of *Neisseria meningitidis* and *N. gonorrhoeae* which encode  
 CC antigenic proteins (see AAY38499-Y38944). The antigenic proteins, their  
 CC fragments, their nucleic acids and antibodies are used for diagnosis,  
 CC prevention (as vaccines) or treatment of *Neisseria* infections,  
 CC such as meningitis, septicaemia and gonorrhea. Both organisms  
 CC are closely related. Fragments of the nucleic acids are useful  
 CC as hybridisation probes and antisense reagents.  
 XX  
 SQ Sequence 1575 BP; 385 A; 387 C; 392 G; 411 T; 0 other;

Query Match 43.9%; Score 100.6; DB 20; Length 1575;  
 Best Local Similarity 96.3%; Pred. No. 8.8e-24;  
 Matches 103; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CGTTTCAGGTTGTCGCGAAGGCTCGGTAAACGGCAACCTGATTACGGGTGATCGAGGCA 60  
 DB 1469 CGTTTCAGGTTGTCGCGAAGGCTCGGTAAACGGCAACCTGATTACGGGTGATCGAGGCA 1528

QY 61 GCTTGAACTTCGCGACGCGAAGCGGAATATGTTATCCCAATGA 107  
 DB 1529 GCTTGAACTTCGCGAAGCGGAATATGTTATCCCAATGA 1575

## RESULT 13

AA15323  
 ID AA15323 standard; DNA; 1575 BP.  
 XX  
 AC AA15323;  
 XX  
 DT 04-SEP-2000 (first entry)  
 XX  
 DE DNA encoding a polypeptide of a Neisseria pathogenic strain.  
 XX  
 KW Pathogenic strain; Neisseria; vaccine; Neisseria infection; ss.  
 OS Neisseria gonorrhoeae.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1..1575  
 FT /\*tag= a  
 XX  
 PN WO200026375-A2.  
 XX  
 PD 11-MAY-2000.  
 XX  
 PF 28-OCT-1999; 99WO-FR02643.  
 XX  
 PR 30-OCT-1998; 98FR-0013693.  
 XX  
 PA (INRM ) PASTEUR MERIEUX SERUMS & VACCINS SA.  
 PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.  
 XX  
 PI Aujame L, Bouchardon A, Renauld-Mongenie G, Rokbi B, Nassif X;  
 PI Tinsley C, Perrin A;  
 DR WPI; 2000-365622/31.  
 DR P-PSDB; AAY93293.  
 XX  
 PT New polypeptide specific for pathogenic Neisseria useful in therapeutic  
 PI or preventative vaccines and for diagnosis -  
 XX  
 PS Claim 4; Page 137-139; 187pp; French.

XX The present sequence encodes a protein that is specific for pathogenic  
 CC strains of Neisseria. The polynucleotides, polypeptides, or their  
 CC antigenic fragments, are used in vaccines to treat or protect against  
 CC Neisseria infections, particularly by N. meningitidis. the  
 CC polynucleotide sequence is also used for recombinant production of  
 CC the polypeptide and to produce attenuated Neisseria strains that  
 CC overexpress it, or express it in a non-toxic mutant form.  
 XX  
 SQ Sequence 1575 BP; 385 A; 387 C; 392 G; 411 T; 0 other;  
 Query Match 43.9%; Score 100.6; DB 21; Length 1575;  
 Best Local Similarity 96.3%; Pred. No. 8.8e-24;  
 Matches 103; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CGGTTTCAGGTTGTCGCGAAGGCTCGGTAACGGCAACCTGATTACGGGTGATCGAGCA 60  
 DB 1469 CGGTTTCAGGTTGTCGCGAAGGCTCGGTAACGAGCAACCTGATTACGGGTGATCGAGCA 1528

QY 61 GCTTGAACTTCGCGAAGCGGAATATGTTATCCCAATGA 107  
 DB 1529 GCTTGAACTTCGCGAAGCGGAATATGTTATCCCAATGA 1575

## RESULT 14

AB241628  
 ID AB241628 standard; DNA; 1467 BP.

AC AB241628;  
 XX  
 DT 07-MAR-2003 (first entry)  
 XX  
 DE N. gonorrhoeae nucleotide sequence SEQ ID 7845.  
 XX  
 KW Antibacterial; infection; vaccine; gene therapy; gene; ds.  
 XX  
 OS Neisseria gonorrhoeae.  
 XX  
 PN WO200279243-A2.  
 XX  
 PD 10-OCT-2002.  
 XX  
 PF 12-FEB-2002; 2002WO-IB02069.  
 XX  
 PR 12-FEB-2001; 2001GB-0003424.  
 XX  
 PA (CHIR-) CHIRON SPA.  
 XX  
 PI Fontana MR, Pizza M, Masignani V, Monaci E;  
 XX  
 DR WPI; 2003-058415/05.  
 DR P-PSDB; ABP80658.

XX New protein from Neisseria gonorrhoeae, useful for the manufacture of a  
 PT medicament for treating or preventing N. gonorrhoeae infection -  
 XX  
 PS Disclosure; Page 760-761; 815pp; English.  
 XX  
 CC The present invention relates to proteins from Neisseria gonorrhoeae.  
 CC Also disclosed are the nucleic acid molecules encoding the proteins and  
 CC antibodies that specifically bind to the proteins. The composition  
 CC comprising the protein, nucleic acid or antibody is useful for the  
 CC manufacture of a medicament for treating or preventing N. gonorrhoeae  
 CC infection, this may be in the form of a vaccine or gene therapy.  
 CC Sequences given in records AB237706-AB242016 represent nucleic acid  
 CC molecules of the invention.  
 XX  
 SQ Sequence 1467 BP; 356 A; 358 C; 377 G; 376 T; 0 other;

Query Match 43.3%; Score 99.2; DB 25; Length 1467;  
 Best Local Similarity 97.1%; Pred. No. 2.6e-23;  
 Matches 101; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CGGTTTCAGGTTGTCGCGAAGGCTCGGTAACGGCAACCTGATTACGGGTGATCGAGCA 60  
 DB 1364 CGGTTTCAGGTTGTCGCGAAGGCTCGGTAACGAGCAACCTGATTACGGGTGATCGAGCA 1423

QY 61 GCTTGAACTTCGCGAAGCGGAATATGTTATCCCAATGA 104  
 DB 1424 GCTTGAACTTCGCGAAGCGGAATATGTTATCCCAATGA 1467

## RESULT 15

AB241663  
 ID AB241663 standard; DNA; 3228 BP.

XX  
 AC AB241663;  
 XX  
 DT 07-MAR-2003 (first entry)  
 XX  
 DE N. gonorrhoeae nucleotide sequence SEQ ID 7915.

XX Antibacterial; infection; vaccine; gene therapy; gene; ds.  
 XX  
 OS Neisseria gonorrhoeae.  
 XX  
 PN WO200279243-A2.  
 XX  
 PD 10-OCT-2002.  
 XX  
 PF 12-FEB-2002; 2002WO-IB02069.

Search completed: November 15, 2003, 00:35:30  
Job time : 150.797 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 14, 2003, 23:58:39 ; Search time 1527.68 Seconds  
(without alignments)  
3643.257 Million cell updates/sec

Title: US-09-928-457-79

Perfect score: 229

Sequence: 1 CGGTTTCAGGTTGCGCAA.....TGGATGCAGCATCAATCCG 229

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

- 1: em\_estba.\*
- 2: em\_esthum.\*
- 3: em\_estin.\*
- 4: em\_estmus.\*
- 5: em\_estov.\*
- 6: em\_estpl.\*
- 7: em\_estro.\*
- 8: em\_hic.\*
- 9: gb\_est1.\*
- 10: gb\_est2.\*
- 11: gb\_hic.\*
- 12: gb\_est3.\*
- 13: gb\_est4.\*
- 14: gb\_est5.\*
- 15: em\_estfun.\*
- 16: em\_estom.\*
- 17: em\_gss\_hum.\*
- 18: em\_gss\_inv.\*
- 19: em\_gss\_pln.\*
- 20: em\_gss\_vrt.\*
- 21: em\_gss\_fun.\*
- 22: em\_gss\_mam.\*
- 23: em\_gss\_mus.\*
- 24: em\_gss\_pro.\*
- 25: em\_gss\_rod.\*
- 26: em\_gss\_phg.\*
- 27: em\_gss\_vrl.\*
- 28: gb\_gss1.\*
- 29: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	40.4	17.6	1201	13 BX417776	BX417776 BX417776
2	37.2	16.2	281	10 BF870084	BF870084 IL3-ET011
3	35.8	15.6	1474	13 BQ073621	BQ073621 AGENCOURT
4	35.6	15.5	707	29 BZ719643	BZ719643 PUCF032TD

5	35	15.3	1012	13 BX338876	BX338876 BX338876
6	35	15.3	1201	13 BX356738	BX356738 BX356738
7	34.8	15.2	841	13 BQ792412	BQ792412 EST_8867
8	34.6	15.1	1397	10 BE908439	BE908439 601503070
9	34	14.8	1005	13 BX382908	BX382908 BX382908
10	33.8	14.8	659	28 AO621897	AO621897 HS_3107_B
11	33.6	14.7	809	13 CB329268	CB329268 IMAGE_1C
12	33.6	14.7	1006	13 BX357256	BX357256 BX357256
13	33.4	14.6	514	10 BE239977	BE239977 EST104026
14	33.4	14.6	670	12 BM765283	BM765283 K-EST0046
15	33.4	14.6	693	12 B1090806	B1090806 602855332
16	33.4	14.6	743	28 AO916226	AO916226 rbe0063A
17	33.4	14.6	777	12 B1087286	B1087286 602850937
18	33.4	14.6	937	28 BH137534	BH137534 ENTNA83TF
19	33.4	14.6	1201	13 BX418945	BX418945 BX418945
20	33.2	14.5	484	12 BM873525	BM873525 laa2h08
21	33	14.4	710	28 BR975720	BR975720 och85a05
22	33	14.4	1201	13 BX422257	BX422257 BX422257
23	32.8	14.3	943	28 AZ206473	AZ206473 SP_0104_B
24	32.6	14.2	525	28 AZ017263	AZ017263 RPCI-23-2
25	32.6	14.2	636	10 BG623921	BG623921 602649079
26	32.4	14.1	429	28 A2499482	A2499482 LM0337H14
27	32.4	14.1	445	9 AA920407	AA920407 LD24006.5
28	32.4	14.1	446	9 AA820402	AA820402 LD23994.5
29	32.4	14.1	632	9 A1542901	A1542901 SD09414.5
30	32.4	14.1	682	10 BG637449	BG637449 SD15454.5
31	32.4	14.1	885	13 BQ959338	BQ959338 AGENCOURT
32	32.4	14.1	974	13 BX344914	BX344914 BX344914
33	32.4	14.1	1101	29 CNS00218	AL061325 Drosophila
34	32.2	14.1	241	12 B1595134	B1595134 Hc_599.07
35	32.2	14.1	244	9 AV282672	AV282672 AV282672
36	32.2	14.1	484	14 CB727453	CB727453 AMGNNUC.M
37	32.2	14.1	669	29 CC176763	CC176763 ZMMBCC030
38	32.2	14.1	670	12 BG773270	BG773270 602722033
39	32.2	14.1	999	10 BE880396	BE880396 601490876
40	32	14.0	1044	13 BX436571	BX436571 BX436571
41	32	14.0	1101	29 CNS0081N	AL108773 Drosophila
42	31.8	13.9	517	12 BM794419	BM794419 K-EST0075
43	31.8	13.9	507	9 AL882926	AL882926 AL882926
44	31.8	13.9	577	29 CC126610	CC126610 ND1.12C23
45	31.8	13.9	594	14 CA048887	CA048887 ssalkh000

#### ALIGNMENTS

RESULT 1  
BX417776  
LOCUS BX417776 Homo sapiens PLACENTA Homo sapiens CDNA clone CS0DE012YN05  
DEFINITION S-PRIME, mRNA sequence.  
ACCESSION BX417776  
VERSION BX417776.1 GI:30654459  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1201)  
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished  
COMMENT Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. Contact: Feng Liang Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/InvitroGen Corporation 1600  
Faraday Avenue Genoscope sequence ID : CS0DE012CG03QPl.  
Location/Qualifiers  
i. .1201  
/organism="Homo sapiens"

```

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DE012YN05"
/tissue_type="PLACENTA"
/clone_lib="Homo sapiens PLACENTA"
/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched, into
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized."
BASE COUNT      171 a      216 c      299 g      268 t
ORIGIN
Query Match      17.6%; Score 40.4; DB 13; Length 1201;
Best Local Similarity 33.1%; Pred. No. 3.2;
Matches 53; Conservative 43; Mismatches 64; Indels 0; Gaps 0;
QY      6 TCAGTGTTCGCGAAGGCTCGGTACCGGCAACCTGATTACGGGTGATGAGCGAGCTTG 65
Db      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      66 AACATTCGCGACGCGAAGCGGATATGTTTATCCGCAATGAGTGGCGTAAACCAATA 125
Db      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      126 AAGCAAAATTTAGATGATGTCGGGGAAGATGCCGACCGA 165
Db      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      427 MGAMTGMAGMTGMGMGGGAGTATGTTGMMGMGMGTGMGGGGGAGMAGMAAMTM 486
Db      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      487 TGMAGMMTGMAGMGMGGGCAAGCGMGTGTMAGGA 526
Db      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 2
LOCUS      BF870084      281 bp      mRNA      linear      EST 17-JAN-2001
DEFINITION      IL3-ET0116-281000-309-B12 ET0116 Homo sapiens cDNA, mRNA sequence.
ACCESSION      BF870084
VERSION      BF870084.1 GI:12260214
KEYWORDS
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL      Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE      20202663
PUBMED      10737800
COMMENT      Contact: Simpson A J G
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/Gethem12.pl?tl=IL3&t2=IL3-ET0116-
281000-309-B12&t3=2000-10-28&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 26
High quality sequence stop: 205.
FEATURES
location/Qualifiers
1..281
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"

```

```

/dev_stage="Adult"
/clone_lib="ET0116"
/note="Organ: lung_tumor; Vector: puc18; Site1: SmaI;
Site 2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT      94 a      48 c      53 g      86 t
ORIGIN
Query Match      16.2%; Score 37.2; DB 10; Length 281;
Best Local Similarity 56.6%; Pred. No. 22;
Matches 69; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
QY      76 ACGCAAGCGGCAATATGTTTATCCGCAATGAGTGGGTAAACCAATAAGACAATT 135
Db      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      136 TAGATGATGTCGGGGAAGATGCCGACCGACAAGACTATGCAAAATATGAAAAACCAAGT 195
Db      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      73 TCGGATATGTCGGGTAGATATTTCTAAACCTCAATATCTGTAAATCGGAAAAACAAA 132
Db      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      196 AC 197
Db      133 AC 134

```

```

RESULT 3
LOCUS      BQ073621/c      1474 bp      mRNA      linear      EST 02-APR-2002
DEFINITION      AGENCOURT 7046430 NIH_MGC_101 Homo sapiens cDNA clone IMAGE:5806473
5', mRNA sequence.
ACCESSION      BQ073621
VERSION      BQ073621.1 GI:19902667
KEYWORDS      EST.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1474)
NIH-MGC http://mgs.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2047 row: f column: 10
High quality sequence stop: 127.
FEATURES
location/Qualifiers
1..1474
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5806473"
/tissue_type="epidermoid carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_101"
/note="Organ: lung; Vector: pOTF7; Site1: EcoRI; Site 2:
XhoI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCGAG(G). Library constructed by Ling Hong in the
Laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."

```



```

BASE COUNT      83 a   423 c   69 g   897 t   2 others
ORIGIN

Query Match      15.6%; Score 35.8; DB 13; Length 1474;
Best Local Similarity 52.8%; Pred. No. 49;
Matches 76; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY      82 AGCGCGAATGTTTATCGGCAATGAGTGGGTAAACCAATTAAGCAAAATTTAGATG 141
        |||||
DB      1248 AAGCGGAAAAGACAAAGAGGGAAGAACAAATATAGAGGAGATAAAGAGAGAAAGAGA 1189
        |||||

QY      142 ATGTGCGGGAAGATCCCGACCGACAGACTATGCAAAATATGAAAAACCAAGTACGGG 201
        |||||
DB      1198 GGAACGGGNAAGAGGAGAGACAGACCAATTAATCGGAAAGGTAGAGACAA 1129
        |||||

QY      202 ATCAGGCGATGATGCACCATCCAA 225
        |||||
DB      1128 AGAAGAGAGGGAAGAGGAGAGAAA 1105
        |||||

RESULT 4
LOCUS      BZ719643/c
DEFINITION PUCFO32D ZM.0.6.1.0 KB Zea mays genomic clone ZMBETA135E15,
            genomic survey sequence.
ACCESSION      BZ719643
VERSION      BZ719643.1 GI:28508337
KEYWORDS      GSS
SOURCE      Zea mays
ORGANISM      Zea mays
REFERENCE      1 (bases 1 to 707)
AUTHORS      Whitelaw,C.A., Quackenbush,J., Van Aken,S., Uterback,T., Resnick
            A., Fraser,C.M., Yuan,F., San Miguel,P., Ma,J. and Bennetzen,J.
TITLE      Maize Genomics Consortium
JOURNAL      Unpublished
COMMENT      Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source           1..707
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMBETA135E15"
                     /clone_lib="ZM.0.6.1.0 KB"
                     /note="Vector: pCR4-TOP0; Site 1: EcoRI; 0.6-1.0 kb high
                     Cor selected genomic DNA library"

BASE COUNT      212 a   175 c   92 g   228 t
ORIGIN

Query Match      15.5%; Score 35.6; DB 29; Length 707;
Best Local Similarity 51.9%; Pred. No. 56;
Matches 80; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY      52 ATGCAGGCGTTGAAACATTCGACGCAAGCGGAATATGTTATCCGCAATGATGG 111
        |||||
DB      532 AGGCAGGACTTGTGGATATTGAGGAAGAAATTGAGGACGATGTTGAGTCAGATTTTAA 473
        |||||

QY      112 CGTAAACCAATTAAGACAAATTTAGATGATGTCGGGAGATGCCCGCCACAGAC 171
        |||||
DB      472 AAGTAGCCCAATTAATGTTGAATATGTTGATACAGTTTCAGATGATGACATGATTC 413
        |||||

QY      172 TATGCAAAATATGAAAAACCAAGTACGGGATCA 205
        |||||

```

---

```

Db      412 TATCGGAGTAATGAAGTTCGCAAGTCTCTTTGATAA 379
        |||||

RESULT 5
LOCUS      BX338876
DEFINITION BX338876 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CSOD1065YA17 5-PRIME, mRNA sequence.
ACCESSION      BX338876
VERSION      BX338876.1 GI:30343727
KEYWORDS      EST.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
REFERENCE      1 (bases 1 to 1012)
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            Li,W.B., Gruber,C., Jesse,J. and Polayes,D.
TITLE      Full-length cDNA libraries and normalization
JOURNAL      Unpublished
COMMENT      Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 4608.f For
            more information about this cluster, see
            http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CSOD1065AA09P1&cluster=4608.f. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CSOD1065AA09P1.
FEATURES             Location/Qualifiers
     source           1..1012
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CSOD1065YA17"
                     /tissue_type="PLACENTA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."

BASE COUNT      295 a   261 c   185 g   160 t   111 others
ORIGIN

Query Match      15.3%; Score 35; DB 13; Length 1012;
Best Local Similarity 39.2%; Pred. No. 79;
Matches 74; Conservative 25; Mismatches 90; Indels 0; Gaps 0;

QY      6 TCAGGTGTGCGGAAGGCTCGGTAAACGGGCAACCTGATTACGGGTGATCGAGGCGAGCTTG 65
        |||||
DB      381 TCCGCGGCTGCGGCTGTTGCCCGATGTGCGGCMTGTGTCGCGCAGMTGAGGCTGG 440
        |||||

QY      66 ACATTCGCGGCGCAAGCGGAATATGTTTATCCGCAATGATGCGGCGTAACCAATA 125
        |||||
DB      441 MGAATGAAAGMTGCMGMAGGGAATAAGTTGAGMGMTGMGGGGGAGMAGMAAATA 500
        |||||

QY      126 AAGACAAATTTAGATGATGTCGGGAAGATGCCCGACGACAGACTATGCAAAATATCA 185
        |||||
DB      501 TGCACGCMWMAAAGMGCMAAGCGCATGMTGAAGSASTATSAASWATAASMTAAAA 560
        |||||

QY      186 AAAACCAAG 194
        |||||
DB      561 ATSMTSMAG 569
        |||||

RESULT 6
LOCUS      BX356738
DEFINITION BX356738 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CSOD1016YD17 5-PRIME, mRNA sequence.
ACCESSION      BX356738

```

```

VERSION      BX356738.1  GI:30366055
KEYWORDS     EST.
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens

REFERENCE
AUTHORS      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE        Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL      1 (bases 1 to 1201)
COMMENT      Full-length cDNA libraries and normalization
Unpublished
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 4608.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0D1016CB09QF1&cluster=4608.f. Contact :
Li.W.B., Gruber C., Jesse, J. and Polayes, D.
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Fazaday Avenue Genoscope sequence ID : CS0D1016CB09QF1.
FEATURES
Location/Qualifiers
1..1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue="CS0D1016YD17"
/clone_lib="PLACENTA.COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with NotI and cloned into the NotI and EcoRV
sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT  414 a 104 c 311 g 265 t 107 others
ORIGIN
Query Match      15.3%; Score 35; DB 13; Length 1201;
Best Local Similarity 50.3%; Pred. No. 79;
Matches 83; Conservative 1; Mismatches 81; Indels 0; Gaps 0;

QY  30  ACGGCGACCTGATTACGGGTGATCGGCGAGGAGCTTGAACATTCGCGAGCGCAAGCGGAA 89
Db  399  ATGTGAGAAATGAATATGGGAGAGAGAGTGGGTGGAGATGAAGATGAGAAAGGGGAA 458
QY  90  TATGTTTATCCGATGATGCGGTGATAAACCATAAGCAATTTAGATGATGTCGGG 149
Db  459  TAAGTTGAAGAGAGATGAAGGGGAGAGAAAATATGAAGAAATAAAAGAGGAAAAG 518
QY  150  GAAGATGCCCGACGACGAAGACTATGCAAAATATGAAAACCAAG 194
Db  519  GAGATGAGTGAAGAGATGATGAAGATAAGATAAATATGTAAR 563

RESULT 7
LOCUS      BQ792412/c
DEFINITION vitifera Grape berries Lambda Triplex2 Library Vitis
ACCESSION  BQ792412
VERSION    BQ792412.1  GI:22007378
KEYWORDS   Vitis vinifera
SOURCE     Vitis vinifera
ORGANISM   Vitis vinifera

REFERENCE
AUTHORS     Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
TITLE       Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; Vitaceae; Vitis.
            1 (bases 1 to 841)
            Abbal, P., Agasse, A., Ageorges, A., Atanassova, R., Barrieu, F.,
            Couture, C., Dedaldechamp, F., Delrot, S., Glissant, D., Grimpiet, J.,
            Hamdi, S., Romieu, C. and Terrier, N.
            Generation of Expressed Sequence Tag from Grape Berry (skin, pulp
            or seeds) at Various Developmental Stages

JOURNAL     Unpublished
COMMENT      Contact: Hamdi S.
UMR 619 - Equipe Biologie de la Vigne
Universite de Bordeaux I, Institut National de la Recherche
Agronomique
71, Avenue Edouard Bourleaux, BP 81, 33883 Villenave D'Ornon Cedex,
France
Tel: 00-33-(0)5-57-12-25-50
Fax: 00-33-(0)5-57-12-25-48
Email: s.hamdi@bordeaux.inra.fr
Seq primer: T7.
FEATURES
Location/Qualifiers
1..841
/organism="Vitis vinifera"
/mol_type="mRNA"
/cultivar="Cabernet Sauvignon"
/db_xref="taxon:29760"
/clone="CT006G02"
/dev_stage="veraison stage"
/clone_lib="Veraison Grape berries Lambda Triplex2
Library"
/note="Organ: Fruit without seeds; Vector: Lambda Triplex2
; Site 1: SfiIA; Site 2: SfiIB; Oriented library"
BASE COUNT  157 a 225 c 116 g 343 t
ORIGIN
Query Match      15.2%; Score 34.8; DB 13; Length 841;
Best Local Similarity 54.8%; Pred. No. 89;
Matches 69; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY  70  TTTCGCGACGCGAAGCGGAATATGTTTATCGCAATGATGCGGTAAACCAATAAAGA 129
Db  475  TTGGCTCTGACAAAGCTAAACCCACAAAGGAAGCATTTGACAAATAGGAAAATA 416
QY  130  CAATTTAGATGATGTCGGGAGAGATGCCGACCGACAGAGACTATGCAAAATATGAAAA 189
Db  415  GAAAGCATGTTGTTGGCAGGATGATTTTAGTGAAAAAGAAATAAGGAAAAAGAGAGAA 356
QY  190  CCAAGT 195
Db  355  CCAAGT 350

RESULT 8
LOCUS      BE908439
DEFINITION BE908439.1  GI:10403019
ACCESSION  BE908439
VERSION    BE908439.1
KEYWORDS   Homo sapiens (human)
SOURCE     Homo sapiens
ORGANISM   Homo sapiens

REFERENCE
AUTHORS     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE       Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL     Unpublished
COMMENT      Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Place: LLN9711 row: k column: 17
High quality sequence scop: 55.
Location/Qualifiers
1..1397
/organism="Homo sapiens"

```

```

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3905032"
/tissue_type="epithelioid carcinoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NH MGC 70"
/notes="Organ: pancreas; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.1 kb. Library constructed by Life
Technologies."
BASE COUNT      655 a      285 c      161 g      296 t
ORIGIN
Query Match      15.1%; Score 34.6; DB 10; Length 1397;
Best Local Similarity 56.6%; Pred. No. 99;
Matches 64; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 114 TAAAAACCAATAAGACAAATTTAGATGATCGGGGAAGATGCCCGACGACAAAGCTA 173
      |||||
Db 604 TATAACAAATAATATAAAATAAAGATATCTCAACATCCCAACAAACACAACTTC 663

QY 174 TGCATAATATGAACCAAGTACGCGATCAGCATGGATGGATGCCATCCCAAT 226
      |||||
Db 664 TTTCACCTAGCTCAACACCGCAGCAGCATGATTTGTATACAGTATCTAAT 716

RESULT 9
BX382908
LOCUS      BX382908      1005 bp      mRNA      linear      EST 08-MAY-2003
DEFINITION Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED
ACCESSION      BX382908
VERSION      BX382908.1
KEYWORDS      EST.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
REFERENCE      1 (bases 1 to 1005)
AUTHORS      Li.W.B., Gruber.C., Jessee.J. and Polayes.D.
TITLE      Full-length cDNA libraries and normalization
JOURNAL      Unpublished
COMMENT      Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 3223.f,
Contact: Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ invitrogen Corporation 1600
Paraday Avenue Genoscope sequence ID : CS0DJ002BB04QF1.
Location/Qualifiers
1. .1005
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DJ002YC08"
/cell_type="T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"
/cell_line="JURKAT"
/clone_lib="Homo sapiens T CELLS (JURKAT CELL LINE) COT
10-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT      380 a      175 c      244 g      195 t      11 others
ORIGIN
Query Match      14.8%; Score 34; DB 13; Length 1005;
Best Local Similarity 50.7%; Pred. No. 1.4e+02;
Matches 76; Conservative 2; Mismatches 72; Indels 0; Gaps 0;

QY 72 CGGACGCGAAGCGGGAATATGTTTATCCGCAATGAGTGGCTAAACCAATAAGACA 131

```

```

Db 414 CKGGGGRGAAGAAGAAAATTTTCAAGACCTCAAAATGATGAAAATCCATTAGATGAC 473
QY 132 AATTATAGATGATGTCGGGAAGATGCCCGACGACAAAGACTATGCAAAATATGAAAAACC 191
      |||||
Db 474 AATTCTGAGGAAGAAATGAAGATGCACCAAAACAAAAGCTTTCTTAAAAAACAAGAAAA 533

QY 192 AAGTACGCGGATCAGGCATGGATGCACGAT 221
      |||||
Db 534 AAGAAACAGAAACCCAGCACAGAAATTATGAT 563

RESULT 10
AQ621897/c
LOCUS      AQ621897/c      659 bp      DNA      linear      GSS 16-JUN-1999
DEFINITION HS-3107_B1_H07_MR CIT Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate=3107 Col=13 Row=P, genomic survey
sequence.
ACCESSION      AQ621897
VERSION      AQ621897
KEYWORDS      GSS.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
REFERENCE      1 (bases 1 to 659)
AUTHORS      Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE      Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL      Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
MEDLINE      99380589
PUBMED      10449764
COMMENT      Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end web server: http://www.htsc.washington.edu
Plate: 3107 row: P column: 13
Seq primer: M13 Reverse
Class: BAC ends
High quality sequence stop: 659.
Location/Qualifiers
1. .859
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="Plate=3107 Col=13 Row=P"
/sex="male"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/notes="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT      104 a      136 c      69 g      350 t
ORIGIN
Query Match      14.8%; Score 33.8; DB 28; Length 659;
Best Local Similarity 64.9%; Pred. No. 1.6e+02;
Matches 50; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 113 GTAAAAACCAATAAGACAAATTTAGATGATCGGGGAAGATGCCCGACGACAAAGACT 172
      |||||
Db 571 GTAAGAAAAATTAAGAAAAATTTAAGGAAGAGGATTAAGCCCAACAAAGAAAGTT 512

QY 173 ATGCAAAATATGAAAAA 189
      |||||
Db 511 AAGCAGACAAAGAAAAA 495

```

```

RESULT 11
CB329268/c 809 bp mRNA linear EST 06-MAR-2003
LOCUS
DEFINITION IMAGE id 4614232 3'end NIH_MGC_92 Homo sapiens cDNA clone
IMAGE:4614232 3', mRNA sequence.
ACCESSION CB329268
VERSION CB329268.1 GI:28873252
KEYWORDS EST.
SOURCE
ORGANISM Homo sapiens (human)
REFERENCE
AUTHORS Okajima,K.
TITLE Unpublished, Kazuki Okajima
JOURNAL Unpublished
COMMENT Contact: Kazuki Okajima
Case Western Reserve University, Genetics
2109 Adelbert Rd, Cleveland, OH 44106, USA
Tel: 216-368-5847
Fax: 216-368-5857
Email: kxo23@po.cwru.edu
Seq primer: M13rev GGAAACAGCATGACCATG
High quality sequence stop: 500
POLYA=Yes
FEATURES
source
LOCATION/Qualifiers
1..809
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4614232"
/tissue_type="embryonal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_92"
/note="Organ: testis; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 2.5 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 184 a 215 c 152 g 242 t 16 others
ORIGIN
Query Match 14.7%; Score 33.6; DB 14; Length 809;
Best Local Similarity 59.4%; Pred. No. 1.8e+02;
Matches 57; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 53 TGCAGGCGAGCTTGACATTCGCGCGGCGGCGGAATGTTTATCCGCAATGAGTGGC 112
|||||
D5 158 TGCATGCGAGTTGTACATTAGTCATTAAAGATGAATTTAGTGTCTTAAAGATTGAA 99
|||||
QY 113 GTAAAACCAATAAGACAAATTTAGATGATGTCGG 148
|||||
D5 98 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAACATGTCGG 63

RESULT 12
BX357256/c 1006 bp mRNA linear EST 05-MAY-2003
LOCUS
DEFINITION BX357256 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
clone CSODI023Y014 5-PRIME, mRNA sequence.
ACCESSION BX357256
VERSION BX357256.1 GI:30382145
KEYWORDS EST.
SOURCE
ORGANISM Homo sapiens (human)
REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 1006)
JOURNAL Full-length cDNA libraries and normalization
COMMENT Unpublished
Contact: Genoscope
Genoscope - Centre National de Sequencage

```

```

BP 191 91006 EVRY cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ InvitroGen Corporation 1600
Paraday Avenue Genoscope, sequence ID : CSODI023BH07Q1.
Location/Qualifiers
1..1006
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSODI023Y014"
/tissue_type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo (dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT 220 a 175 c 187 g 221 t 203 others
ORIGIN
Query Match 14.7%; Score 33.6; DB 13; Length 1006;
Best Local Similarity 33.5%; Pred. No. 1.8e+02;
Matches 57; Conservative 37; Mismatches 76; Indels 0; Gaps 0;

QY 6 TCAGTTTCGCGAGGCTCGTAACGGCGCACTGATTACGGGTGATGCGAGGAGCTTG 65
|||||
D5 771 TWSSSCGGGGGGGGGGGAATAAATAAACCCCTAAWGGGKGSGGSRRTTAKAWA 712
|||||
QY 66 AACATTTCGCGAGGCGAAGCGGGAATATGTTTATCCGCAATGAGTGGCGTAAAAACCAATA 125
|||||
D5 711 ARRTWTTRTGGGGGAGCGGCGCCCAATTSSCCSGGCGMCGGARMCAAAWAAAAA 652
|||||
QY 126 AGACAAATTAGATGATGTCGGGAGATGCGCGACCGACGCAAGACTATG 175
|||||
D5 651 AWAARRAATWAAGATAACMGCGGGAARAVRSGSTAAAAAAMTACCCKG 602
|||||

RESULT 13
BX239977 514 bp mRNA linear EST 12-JUL-2000
LOCUS
DEFINITION EST0404026 MHRP- Medicago truncatula cDNA clone pMHRP-41122, mRNA
sequence.
ACCESSION BX239977
VERSION BX239977.1 GI:9056160
KEYWORDS EST.
SOURCE Medicago truncatula (barrel medic)
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.
REFERENCE 1 (bases 1 to 514)
AUTHORS Harrison,M.J., Town,C.D., Bowman,C.L., Craven,M.B., Hansen,T.S.,
Holt,I.E., Cho,J. and Fraser,C.M.
TITLE ESTs from phosphate-starved roots of Medicago truncatula
JOURNAL Unpublished
COMMENT Contact: Maria J. Harrison
Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73401, USA
Tel: 580-223-5810
Fax: 580-221-7380
Email: mjharrison@noble.org
The Samuel Noble Roberts Foundation: N265844e
TIGR sequence name: MTHA59TK
More information is available at
http://chrysis.tamu.edu/medicago
Seq primer: SKmod (CTA GAA CTA g-g GAT CC)
Location/Qualifiers
1..514
/organism="Medicago truncatula"
/mol_type="mRNA"
FEATURES
source

```

```

/cultivar="A17"
/db_xref="taxon:3880"
/clone="pMHRP-41122"
/tissue_type="roots"
/dev_stage="phosphate-starved"
/lab_host="XLOLR"
/clone_lib="MHRP-"
/notes="Vector: pBluescript SK-; Site 1: EcoRI; Site 2: XhoI; At the trifoliate stage, M. truncatula plants were transplanted to phosphate-free sand and grown for a further 30 days. During this period, they were fertilized twice weekly with 1/2 Hoaglands solutions containing 200M potassium phosphate. cDNA was prepared from polyA+ enriched RNA. The cDNA was directionally ligated into the UniZap Xr vector from Strategene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-Zap phage using Ex-assist helper phage and propagated in XLOLR cells."
BASE COUNT      176 a      89 c      103 g      146 t
ORIGIN
Query Match      14.6%; Score 33.4; DB 10; Length 514;
Best Local Similarity 52.5%; Pred. No. 2.1e+02;
Matches 73; Conservative 0; Mismatches 66; Indels 0; Gaps 0;
QY 65 GAACATTCGCGCGCAAGCGGATATCTTTATCCGCAATGAGTGGCGTAAACCAAT 124
Db 313 GTACTTACAGATGAGGATGATACATCTGTAAACCTTTTGGTGGAGGAATACCTTAT 372
QY 125 AAAGACAAATTTAGATGATGTCGGGAAGATGCCGACCGACAGACATGCAAAATATG 184
Db 373 AAAGACACACTTTAAGGCTCTATGATCAAAATTTACTTCTACCATCATGATGATGAC 432
QY 185 AAAACCAAGTACGGGAT 203
Db 433 AAACGAGTGCACCAAT 451

```

```

RESULT 14
BM765283
LOCUS
DEFINITION
X-EST0046949 S7SNU719s1 Homo sapiens cDNA clone S7SNU719s1-3-E09
5', mRNA sequence.
BM765283
BM765283.1 GI:19094898
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Kim, N.S., Hahn, Y., Oh, J.H., Lee, J.Y., Ahn, H.Y., Chu, M.Y., Kim, M.R.,
Oh, K.J., Cheong, J.E., Sohn, H.Y., Kim, J.M., Park, H.S., Kim, S. and
Kim, Y.S.
21C Frontier Korean EST Project 2001
Unpublished
Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 3 row: E column: 09
High quality sequence stop: 670.
Location/Qualifiers
1. 670
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="S7SNU719s1-3-E09"
/sex="M"
FEATURES
source

```

```

/tissue_type="Stomach"
/cell_type="Epithelial"
/cell_line="SNU-719"
/lab_host="T0p10F"
/clone_lib="S7SNU719s1"
/notes="Organ: Stomach; Vector: pCNS; Site 1: EcoRI; Site 2: NotI; The poly (A)+ RNA was dephosphorylated with bacterial alkaline phosphatase (BAP) and then decapped with tobacco acid pyrophosphatase (TAP). The decapped intact mRNA was ligated with DNA-RNA linker including EcoR I site by treatment of T4 RNA ligase and the first strand cDNA was synthesized from oligo dt-selected mRNA by priming with dt-tailed vector. The dt-tailed vector was adjusted to have about 60nt. The cDNA vector was circularized with E. coli DNA ligase after digestion of EcoRI which site is also included in vector. An RNA strand converted to a DNA strand by Okayama-Berg method. The obtained cDNA vectors were used for transformation of competent cells E. coli Top10F by electroporation method. The cDNA libraries constructed by this method are full-length enriched cDNA library. After analyzing and sequencing about 2,000 ~ 3,000 colonies in original cDNA library, the abundant cDNAs were selected and amplified by PCR reaction using vector region primer including T7 promoter as 5' primer and N(OT)14 as 3' primer. The PCR products were used as template for synthesis of biotinylated single stranded RNA by in-vitro transcription reaction. The synthesized RNA probes were hybridized with antisense single stranded cDNAs prepared from original library and incubated with avidin-gel. After removing DNA-RNA hybrids by centrifuge, the substracted cDNA libraries were constructed by transfection of the remaining DNA into competent cells E. coli Top10F with electroporation method."
BASE COUNT      274 a      111 c      160 g      125 t
ORIGIN
Query Match      14.6%; Score 33.4; DB 12; Length 670;
Best Local Similarity 51.7%; Pred. No. 2.1e+02;
Matches 76; Conservative 0; Mismatches 71; Indels 0; Gaps 0;
QY 75 GACGCAAGCGGAATATGTTTATCCGCAATGAGTGGCGTAAACCAATAGACAAAT 134
Db 248 GAGGAGGAAGAGAAAATTTTCAAGACCTCAATTTGATGAAATCCATTAGTACAAAT 307
QY 135 TTAGATGATGTCGGGAAGATGCCGACCGACAGACTATGCAAAATATGAAAAACAAG 194
Db 308 TCTGAGGAAGAAATGGAAGATGCACCAACAAAAGCTTTCTAAAAACACAGAAAAAG 367
QY 195 TAGCGCGATCAGGCATGGATGCACGAT 221
Db 368 AAACAGAAACACACAGAAATTATGAT 394
RESULT 15
BI090806
LOCUS
DEFINITION
602855332F1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:4996855 5',
RNA sequence.
BI090806
BI090806.1 GI:14509136
EST.
KEYWORDS
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-r@mail.nih.gov
Tissue Procurement: ATCC

```

cdNA Library Preparation: Life Technologies, Inc.  
cdNA Library Arrayed by: Incyte Genomics, Inc.  
DNA sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:

http://image.iitl.gov  
Plate: LLAM11022 row: p column: 08  
High quality sequence stop: 680.

## FEATURES

source

```

1. 523
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4996855"
/cell_line="MGC36"
/lab_host="DH10B"
/clone_lib="NH MGC 10"
/notes="Organ: cervix; Vector: pCMV-SPOPT6; Site.1: NotI; Site.2: SalI; Cloned unidirectionally. Primer: Oligo dr. Average insert size 1.5 kb. Library prepared by Life Technologies"

```

	287 a.	119 c	150 g	137 b
BASE COUNT				
ORIGIN				

Query Match	14.6%	Score 33.4;	DB 12;	Length 693;
Best Local Similarity	51.7%;	Pres. No. 2.1e+02;		
Matches	76;	Conservative	0; Mismatches 71;	Indels 0; Gaps 0;
75 GACGGCAAGGGGAATATGTTTTATCCGCAATCAGTGGCGTAAACCAATTAAGACAAT	134			
98 GAGGAGGAAGAAGAAAATTTTTCAAGACCTCAAAATTGATGAAATTCCTATTAGTACCAAT	157			
135 TTAGATGATTCGGGGGAAGATGCCCGACGCAAGACTATGCAAAATATGAAAAACCAAG	194			
158 TCTGAGGAAGAAATGAGNAGTGACCCAAACAAAGCTTTCTAAAAACACAGAGAAAAG	217			
195 TACCGGATCAGCGATGGATGCAGAT	221			
218 AAACAGAAAACGACACAGAAATTATGAT	244			

Search completed: November 15, 2003, 08:02:31  
Job time : 1530.68 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 15, 2003, 00:06:39 ; Search time 35.6666 Seconds  
(without alignments)  
2833.774 Million cell updates/sec

Title: US-09-928-457-79  
Perfect score: 229  
Sequence: 1 CGGTTTCAGGTTCGCCAA.....TGGATGCAATCAATCG 229

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents, NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	33	14.4	7218	1	US-08-232-463-14
2	31.6	13.8	552	4	US-09-370-838-266
3	31.6	13.8	1583	4	US-09-370-838-224
4	31	13.5	1626	4	US-08-134-001C-1644
5	30.6	13.4	612	4	US-09-107-532A-1427
6	29.8	13.0	1923	4	US-09-620-312D-77
7	29.4	12.8	11887	4	US-08-961-527-146
8	28.4	12.4	289	3	US-09-007-005-17
9	28.4	12.4	289	3	US-09-244-796-17
10	28.4	12.4	3729	4	US-09-107-532A-1587
11	28.2	12.3	700	3	US-08-998-416-302
12	28.2	12.3	714	3	US-08-998-416-863
13	28.2	12.3	714	3	US-08-998-416-1139
14	28.2	12.3	722	3	US-08-998-416-680
15	28.2	12.3	725	3	US-08-998-416-1051
16	28.2	12.3	761	3	US-08-998-416-382
17	28.2	12.3	1512	4	US-08-328-352-1102
18	28	12.2	3713	1	US-08-100-709-1
19	28	12.2	3713	1	US-08-176-865-1
20	28	12.2	3713	1	US-08-474-038-1
21	28	12.2	3713	2	US-08-779-046-1
22	28	12.2	3713	2	US-08-881-340-1
23	27.8	12.1	3501	1	US-08-448-170-5
24	27.8	12.1	3501	3	US-08-961-803-4
25	27.6	12.1	615	4	US-08-134-001C-474
26	27.6	12.1	1664976	4	US-08-916-421B-1
27	27.4	12.0	277	3	US-09-007-005-3

Query Match 14.4%; Score 33; DB 1; Length 7218;

28	27.4	12.0	277	3	US-09-244-796-3	Sequence 3, Appli
29	27	11.8	922	4	US-09-221-017B-692	Sequence 692, App
30	27	11.8	2445	4	US-09-339-159B-25	Sequence 25, Appl
31	27	11.8	3607	1	US-08-647-351B-1	Sequence 1, Appli
32	27	11.8	3607	4	US-09-326-074-1	Sequence 1, Appli
33	27	11.8	3804	4	US-09-620-312D-894	Sequence 894, App
34	27	11.8	72928	3	US-09-009-913-1	Sequence 1, Appli
35	27	11.8	1830121	4	US-09-557-884-1	Sequence 1, Appli
36	27	11.8	1830121	4	US-09-643-990A-1	Sequence 1, Appli
37	26.8	11.7	422	3	US-09-037-990B-12	Sequence 12, Appl
38	26.8	11.7	2277	1	US-08-676-967-2	Sequence 2, Appli
39	26.8	11.7	2277	1	US-08-676-974-2	Sequence 2, Appli
40	26.8	11.7	2277	2	US-09-098-487-2	Sequence 2, Appli
41	26.6	11.6	784	3	US-08-961-083-151	Sequence 151, App
42	26.6	11.6	784	4	US-09-536-784-151	Sequence 151, App
43	26.6	11.6	1289	4	US-09-247-155-138	Sequence 138, App
44	26.6	11.6	4643	4	US-09-453-702B-22	Sequence 22, Appl
45	26.6	11.6	6474	4	US-08-961-527-155	Sequence 155, App

ALIGNMENTS

RESULT 1  
US-08-232-463-14/c  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFELINGER, F.  
; APPLICANT: FALKNER, F. G.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 1800 Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22313-0299

COMPUTER READABLE FORM:  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232,463  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/935,313  
FILING DATE:  
APPLICATION NUMBER: EP 91 114 300.6  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: (703)683-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: pizgpt-F1s  
US-08-232-463-14

	Best Local Similarity	1.5%, Pred. No.	0.17;	Mismatches	76;	Indels	0;	Gaps	0;
	Matches	3;	Conservative	126;					
QY	18	GAAGCTCGGTAAACGGCAACCTGATTACGGTGATCAGGCAGCTTGAACATTCGGCAC	77	:	:	:	:	:	:
Db	1268	RRR	1209	:	:	:	:	:	:
QY	78	GCAGAAGCGGGAATATGTTTATCCGCAATGAGTGGCGTAAAAACCAATAAAGACAATTA	137	:	:	:	:	:	:
Db	1208	RRR	1149	:	:	:	:	:	:
QY	138	GATGATGTCGGGAAGATGCCACCGACAAGCATATGCAAAATATGAAAACCAAGTAC	197	:	:	:	:	:	:
Db	1148	RRR	1089	:	:	:	:	:	:
QY	198	CGCGATCAGGCATGATGCACGATC	222	:	:	:	:	:	:
Db	1088	RRR	1064	:	:	:	:	:	:

```

RESULT 2
US-09-370-838-266
; Sequence 266, Application US/09370838
; Patent No. 644425
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Roadoh
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOUNDS FOR THERAPY AND DIAGNOSIS OF
; TITLE OF INVENTION: LONG-CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.475C1
; CURRENT APPLICATION NUMBER: US/09/370,838
; CURRENT FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/285,323
; EARLIER FILING DATE: 1999-04-02
; NUMBER OF SEQ ID NOS: 289
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 266
; LENGTH: 552
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-370-838-266

```

RESULT 3  
US-09-370-838-224  
: Sequence 224 Application US/09370838  
: Patent No. 6444325  
: GENERAL INFORMATION:  
: APPLICANT: Reed, Steven G.  
: APPLICANT: Lcdes, Michael J.  
: APPLICANT: Mcnamara, Rodolph

```

; APPLICANT: Secrlist, Heather
;
; TITLE OF INVENTION: COMPOUNDS FOR THERAPY AND DIAGNOSIS OF
;
; TITLE OF INVENTION: LUNG CANCER AND METHODS FOR THEIR USE
;
; FILE REFERENCE: 210121.475C1
;
; CURRENT APPLICATION NUMBER: US/09/370,838
;
; CURRENT FILING DATE: 1999-08-09
;
; EARLIER APPLICATION NUMBER: US 09/285,323
;
; EARLIER FILING DATE: 1999-04-02
;
; NUMBER OF SEQ ID NOS: 289
;
; SOFTWARE: FastSEQ for Windows Version 3.0
;
; SEQ ID NO 224
;
; LENGTH: 1583
;
; TYPE: DNA
;
; ORGANISM: Homo sapien
;
; US-09-370-838-224

```

```

RESULT 4
US-09-134-001C-1644
; Sequence 1644, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1644
; LENGTH: 1626
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1644

```

Query Match	13.5%;	Score 31;	DB 4;	Length 1626;
Best Local Similarity	52.8%;	Pred. No. 0.44;		
Matches	67;	Conservative 0;	Mismatches 60;	Indels 0; Gaps 0;
QY	63	TTGACATTCGCGACGGCGGAATATGTTTATCCGCAATGAGTGGCGTAAAAACCA	122	
Db	279	TTTATCATTAGTCAGTACCGGTATAATACTGTGATTGACATGAGTATTGGGATTAATCA	338	
QY	123	ATAAGACAAATTTAGATGATGTCGGGGAAGATGCCCGACCAAGACTATGCAAAATA	182	
Db	339	ATACCACACAGATAAGACAATACACGTAATTTTATAAACCACTCATTTATGAAAAATGA	398	
QY	183	TGAAAAA	189	



Db 399 AGACAA 405

RESULT 5

US-09-107-532A-1427

; Sequence 1427, Application US/09107532A

; Patent No. 6583275

; GENERAL INFORMATION:

; APPLICANT: Lynn A Doucette-Stamm and David Bush

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS

; NUMBER OF SEQUENCES: 7310

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: GENOME THERAPEUTICS CORPORATION

; STREET: 100 Beaver Street

; CITY: Waltham

; STATE: Massachusetts

; COUNTRY: USA

; ZIP: 02354

; COMPUTER READABLE FORM:

; MEDIUM TYPE: CD-ROM ISO9660

; COMPUTER: PC

; OPERATING SYSTEM: <Unknown>

; SOFTWARE: ASCII

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/107,532A

; FILING DATE: 30-Jun-1998

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: 60/085,598

; FILING DATE: 14 May 1998

; APPLICATION NUMBER: 60/051571

; FILING DATE: July 2, 1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Ariniello, Pamela Deneke

; REGISTRATION NUMBER: 40,489

; REFERENCE/DOCKET NUMBER: GTC-012

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (781)893-5007

; TELEFAX: (781)893-8277

; INFORMATION FOR SEQ ID NO: 1427:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 612 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: circular

; MOLECULE TYPE: DNA (Genomic)

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

; ORIGINAL SOURCE:

; ORGANISM: Enterococcus faecium

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (8) LOCATION 1...612

; SEQUENCE DESCRIPTION: SEQ ID NO: 1427:

US-09-107-532A-1427

Query Match 13.4%; Score 30.6; DB 4; Length 612;

Best Local Similarity 60.0%; Pred. No. 0.39;

Matches 51; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 55 CAGCGAGTTGAACATTCGCGGCAAGCGGGAATATGTTATCCGCAATGAGTGGCGT 114

DB 330 CAAGCAAGCAATGATATTTACGGGAAGACTTGAACAGTTTACGGGAATCATTTGGCGA 389

QY 115 AAAAACCATAAGACAAATTAGA 139

DB 390 AAAAGCAAAACACAAATATACA 414

RESULT 6

US-09-620-312D-77

; Sequence 77, Application US/09620312D

; Patent No. 6569662

; GENERAL INFORMATION:

; APPLICANT: Tang, Y. Tom

; APPLICANT: Liu, Chenghua

; APPLICANT: Asundi, Vinod

; APPLICANT: Zhang, Jie

; APPLICANT: Ren, Feiyang

; APPLICANT: Chen, Rui-hong

; APPLICANT: Zhao, Qing A.

; APPLICANT: Wehrman, Tom

; APPLICANT: Xue, Aidong J.

; APPLICANT: Yang, Yonghong

; APPLICANT: Wang, Jian-Rui

; APPLICANT: Zhou, Ping

; APPLICANT: Ma, Yunging

; APPLICANT: Wang, Dunrui

; APPLICANT: Wang, Zhiwei

; APPLICANT: John Tillinghast

; APPLICANT: Dmanac, Radcoje T.

; TITLE OF INVENTION: No. 658662el Nucleic Acids and

; TITLE OF INVENTION: Polypeptides

; FILE REFERENCE: 784CIP2B

; CURRENT APPLICATION NUMBER: US/09/620,312D

; CURRENT FILING DATE: 2000-07-19

; PRIOR APPLICATION NUMBER: 09/552,317

; PRIOR FILING DATE: 2000-04-25

; PRIOR APPLICATION NUMBER: 09/488,725

; PRIOR FILING DATE: 2000-01-21

; NUMBER OF SEQ ID NOS: 1105

; SOFTWARE: pt\_FL\_genes Version 1.0

; SEQ ID NO 77

; LENGTH: 1923

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (112)...(1233)

US-09-620-312D-77

Query Match 13.0%; Score 29.8; DB 4; Length 1923;

Best Local Similarity 49.7%; Pred. No. 1.2;

Matches 76; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 60 AGCTTGAACATTCCGCGACGCAAGCGGGAATATGTTATCCGCAATGAGTGGCGTAAAAA 119

DB 33 AGACAGCACACTGCTGACTGTTTTCAGTTGTTCTGTAACAGCAGAAAGTGCATCTACTA 92

QY 120 CCAATAAAGACAAATTTAGATGATGTCGGGGAAGATGCCGACCCGACAGAAATATGCAAA 179

DB 93 GGAGTAGTCAGAAATCAAAATGCTGAAGAGAAAGCCATCCAATGTTTCAGAGAAGGAGAA 152

QY 180 ATATGAAAACCAAGTACCGCGATCAGGCATCG 212

DB 153 ACATCAAAAACCAAGCGAAGCAGCAGTTTGG 185

RESULT 7

US-08-961-527-146

; Sequence 146, Application US/08961527

; Patent No. 6420135

; GENERAL INFORMATION:

; APPLICANT: Charles Kunsch

; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences

; NUMBER OF SEQUENCES: 391

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue

; CITY: Rockville

; STATE: Maryland

; COUNTRY: USA

; ZIP: 20850

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

; COMPUTER: HP Vectra 486/33

```

, OPERATING SYSTEM: MSDOS version 6.2
,
, SOFTWARE: ASCII Text
,
, CURRENT APPLICATION DATA:
,
, APPLICATION NUMBER: US/08/961.527
,
, FILING DATE:
,
, CLASSIFICATION: 424
,
, PRIOR APPLICATION DATA:
,
, APPLICATION NUMBER:
,
, FILING DATE:
,
, ATTORNEY/AGENT INFORMATION:
,
, NAME: Brookes, A. Anders
,
, REGISTRATION NUMBER: 36,373
,
, REFERENCE/DOCKET NUMBER: PB340P1
,
, TELECOMMUNICATION INFORMATION:
,
, TELEPHONE: (301) 309-8504
,
, TELEFAX: (301) 309-8512
,
, INFORMATION FOR SEQ ID NO: 146:
,
, SEQUENCE CHARACTERISTICS:
,
, LENGTH: 11987 base pairs
,
, TYPE: nucleic acid
,
, STRANDEDNESS: double
,
, TOPOLOGY: linear
,
, US-08-961-527.146

```

	Query Match	12.8%	Score 29.4	DB 4	Length 11887
	Best Local Similarity	52.0%	Pred. No. 3.8		
	Matches	66	Conservative 0	Mismatches 61	Indels 0
					Gaps 0
QY	88	AATATCTTTTATCCGCAATGATGGCTTAAACCAATAAAGACAATTTAGATGATGCG	147		
Db	11498	AACCTTCTAATCTGATACCAATTGGGTAAAGGACCAATTCGTACTCTTAGATGGAACG	11557		
QY	148	GGGAAGATGCCCGACCGACAAGACTATGCAAAATATGAAAAACCAAGTACGGCGATCAGG	207		
Db	11558	GGATATATGCCAGCGCAACCAATCTCTCCAGATAGCAATCAAAATATAGTACTTAC	11617		
QY	208	CATGGAT	214		
Db	11618	CATCAAT	11624		

```

RESULT 8
US-09-007-005-17      Application US/09007005B
; Sequence 17, Application US/09007005B
; Patent No. 6258558
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihe
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; TITLE OF INVENTION: FUSIONS
; FILE REFERENCE: 00786/350003
; CURRENT APPLICATION NUMBER: US/09/007,005B
; CURRENT FILING DATE: 1998-01-14
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064,491
; EARLIER FILING DATE: 1997-11-06
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 289
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
; FEATURE:
; NAME/Key: misc feature
; LOCATION: (1)..(289)
; OTHER INFORMATION: n = A,T,C or G
US-09-007-005-17

```

Query Match  
12.4%; Score 28.4; DB 3; Length 289;

```

Best Local Similarity 15.2%; Pred.No. 1.7;
Matches 25; Conservative 56; Mismatches 83; Indels 0; Gaps 0;

Qy 26 GGTACGGGCAACTGATTACGGGTGATCGAGCAGCTTGAACTTCGACGGCAAGGC 85
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 122 SRNRNSRNRNSRNRNSRNRNSRNRNSRNRNSRNRNSRNRNSRNRNSRNRNSRNR 181
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 86 GGAATATGTTATCCCAATGACTGCGTAAACCAATAAGACAAATTTAGATGATGT 145
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 182 SRNRNSRNRNSRNRNSRNRNSRNRNSRNRNSRNRNSRNRNSRNRNSRNRNSRNR 241
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 146 CGCGGAGATGCCGACCGACAAGACTATGCAAAAATATGAAAA 189
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 242 CRURCURURGRCGRUAAAAAIAAAAAAAAAAAAAAAAAAAAA 285
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 9
US-09-244-796-17
; Sequence 17, Application US/09244796
; Patent No. 6281344
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihe
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; FILE REFERENCE: 00786/350007
; CURRENT APPLICATION NUMBER: US/09/244,796
; CURRENT FILING DATE: 1999-02-05
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064,491
; EARLIER FILING DATE: 1997-11-06
; EARLIER APPLICATION NUMBER: 09/007,005
; EARLIER FILING DATE: 1998-01-14
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 289
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(289)
; OTHER INFORMATION: n = A,T,C or G
;
US-09-244-796-17

```

[illegible]

RESULT 10  
US-09-107-532A-1587  
; Sequence 1587, Application US/09107532A  
; Patent No. 6583275  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm  
; TITLE OF INVENTION: NUCLEIC ACID

```

; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
;
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107.532A
; FILING DATE: 30-Jun-1998
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
;
; INFORMATION FOR SEQ ID NO: 1587:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3729 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (B) LOCATION 1...3729
; SEQUENCE DESCRIPTION: SEQ ID NO: 1587:
;
; US-09-107-532A-1587
;
; Query Match          12.4%; Score 28.4; DB 4; Length 3729;
; Best Local Similarity 56.4%; Pred.No.5.1; Indels 0; Gaps 0;
; Matches 53; Conservative 0; Mismatches 41;
;
; QY      63 TTGAACATTCCGACGGCAGCGCGGAATATGTTTTATCCGAATGAGTGCGGTAAACAACCA.122
;         |||
; Db      1138 TTCAGATTCTCACAGAAAAACGAAAGATGTATGGCTGCCAGTAGAGCTTCAAAACAT 1197
;         |||
;
; QY      123 ATAAAGACAATTAGATGATGTCGGGGAAGATG 156
;         |||
; Db      1198 TATCGAAAAAATTCAAAAGAAGTAATGCTAGATG 1231
;         |||
;
; RESULT 11
; US-08-998-416-302
; Sequence 302, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippsen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jürgen
; APPLICANT: Knechtle, Philipp
; APPLICANT: Rebschung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPILII AND USES THEREOF

```



Search completed: November 15, 2003, 08:09:09  
Job time : 39.6686 secs



## ALIGNMENTS

## RESULT 1

US-09-928-457-79

: sequence /9, Application US/09928457  
: Patent No. US20020164603A1

; GENERAL INFORMATION:

APPLICANT:

;; TITLE OF INVENTION: DNA, specific proteins and peptides

; TITLE OF INVENTION: for

; NUMBER OF SEQUENCES: 99

```

; ; COMPUTER READABLE FORM:
; ; MEDIUM TYPE: Floppy disk

```

COMPUTER: IBM PC compatible

```

;
OPERATING SYSTEM: PC-DOS/MS-DOS

```

SOFTWARE: PATENT IN RELATION TO CURRENT APPLICATION DATA.

APPLICATION NUMBER: US/09/928,457

FILING DATE: 2001-08-14

PRIOR APPLICATION DATA: 09/214-759  
 APPLICATION NUMBER: 09/214-759

FILING DATE: 199-12-10

; INFORMATION FOR SEQ ID NO: 79:

SEQUENCE CHARACTERISTICS:  
LENGTH: 229 base pairs

```

;
; TYPE: nucleotide
;

```

; STRANDEDNESS: single

TOPOLOGY: Linear

HYPOTHETICAL: NO

; ANTI-SENSE:

Query Match . 100.0%; Score 229; DB 10; Length 229;

Best Local Similarity 100.0%; Pred. No. 4.9e-65;

```
Matches 229; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 1 CGGTTTCAGGTTGTCGCGAAGGCTCGGTACCGGCAACCTGATTACGGGTGATGCAGGCA 60

db: 1 CGGTTTCAGGTTGTCCGCGAAGGTCGGTAA CGGGGAA CCTGATTA CGGGTCATGCAGGCA 60

Database : Published\_Applications\_NA:\*

```
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
```

```
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB_seq:*
```

```
3: /cgn2_6/prodata/1/pubpna/us06_NEW_PUB.seq:*
4: /cgn2_6/prodata/1/pubpna/us06_PUBCOMB.seq:*
```

```
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB_seq:*
```

```
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
```

```

7: /cgn2_6/prodata/1/pubna/US08_NEW_PUB.seq:*
8: /cgn2_6/prodata/1/pubna/US08_PUBCOMB.seq:*

```

```
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
```

```
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:
11: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:
12: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:
```

```

11: /cgn2_6/ptodata/t/pub/pna/US09C_PUBCOMP.B.seq:
12: /cgn2_6/ptodata/t/pub/pna/US09_NEW_PUB.seq:

```

13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:

```
14: /cgn2_6/ptodata/1/pubna/US10B_PUBCOMB.seq:
15: /cgn2_6/ptodata/1/pubna/US10B_PUBCOMB.seq:
16: /cgn2_6/ptodata/1/pubna/US10B_PUBCOMB.seq:
```

```

15: /cgn2_6/ptodata/I/pub/pna/USTO_NEW_PUB.seq:
16: /cgn2_6/ptodata/I/pub/pna/USTO_NEW_PUB.seq:

```

17: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:★

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	229	100.0	229	US-09-928-457-79	Sequence 79, Appl
C	162.4	70.9	192	US-09-928-457-67	Sequence 67, Appl
3	33.4	14.6	1828	US-09-764-864-190	Sequence 190, App
4	33.4	14.6	1874	US-09-922-846-188	Sequence 188, App
5	32.4	14.1	3891	US-10-108-605-348	Sequence 348, App
6	31.6	13.8	552	US-09-738-973-266	Sequence 266, App
7	31.6	13.8	552	US-09-854-173-266	Sequence 266, App
8	31.6	13.8	552	US-10-144-649A-266	Sequence 266, App
9	31.6	13.8	1538	US-10-044-090-43	Sequence 43, Appl
10	31.6	13.8	1583	US-09-738-973-224	Sequence 224, App
11	31.6	13.8	1583	US-09-854-133-224	Sequence 224, App
12	31.6	13.8	1583	US-10-144-649A-224	Sequence 224, App
13	31.4	13.7	12100	US-09-764-891-9126	Sequence 9126, App
14	31.4	13.7	12100	US-10-091-572-702	Sequence 702, App
15	31	13.5	1251	US-09-764-864-625	Sequence 625, App
C	30.6	13.4	5657	US-09-974-298-96	Sequence 96, Appl

```

QY 61 GCTTGAACATTCGGACGCGAAGCGGAATATGTTTATCCGCAATGAGTGGCGTAAAAAC 120
DB 61 GCTTGAACATTCGGACGCGAAGCGGAATATGTTTATCCGCAATGAGTGGCGTAAAAAC 120
QY 121 CAATAAAGACAAATTTAGATGATGTCGGGGAAGATGCCGACCGCAAGACTATGCAAAA 180
DB 121 CAATAAAGACAAATTTAGATGATGTCGGGGAAGATGCCGACCGCAAGACTATGCAAAA 180
QY 181 TATGAAAAACCAAGTACCGCGATCAGGATCGATGCGATGCGATCCCAATCCG 229
DB 181 TATGAAAAACCAAGTACCGCGATCAGGATCGATGCGATGCGATCCCAATCCG 229

```

## RESULT 2

```

US-09-928-457-67/c
; Sequence 67, Application US/09928457
; Patent No. US20020164603A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: DNA, specific proteins and peptides
; TITLE OF INVENTION: of the Neisseria meningitidis species bacteria, method
; TITLE OF INVENTION: for obtaining them and their biological application.
; NUMBER OF SEQUENCES: 99
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30 (OEB)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/928,457
; FILING DATE: 2001-08-14
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/214,759
; FILING DATE: 199-12-10
; INFORMATION FOR SEQ ID NO: 67:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 192 base pairs
; TYPE: nucleotide
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-09-928-457-67

```

```

Query Match 70.9%; Score 162.4; DB 10; Length 192;
Best Local Similarity 93.8%; Pred. No. 3.2e-43;
Matches 180; Conservative 0; Mismatches 11; Indels 1; Gaps 1;

QY 14 TCGCAAGGCTCGGTAAACGGCAACCTGATTACGGGTGATGCGAGCAGCTTGAACTTCG 73
DB 192 TCTCGAAGCTCGGTAAACGGCAACCTGATTACGGGTGATGCGAGCAGCTTGAACTTCG 133
QY 74 CGACGGCAAGCGGGAATR-TGTTTATCCGCAATGAGTGGGTAAAAACCAATAAAGCAA 132
DB 132 CGACGGCAAGCGGGAATR-TGTTTATCCGCAATGAGTGGGTAAAAACCAATAAAGCAA 73
QY 133 ATTATGATGATGTCGGGGAAGATCCCGACCGCAAGACTATGCAAAATATGAAAAACCA 192
DB 72 ATTATGATGATGTCGGGGAAGATCCCGACCGCAAGACTATGCAAAATATGAAAAACCA 13
QY 193 AGTACGGGATC 204
DB 12 AGTACGGGATC 1

```

## RESULT 3

```

US-09-764-864-190
; Sequence 190, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.

```

```

; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 190
; LENGTH: 1828
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-864-190

```

```

Query Match 14.6%; Score 33.4; DB 10; Length 1828;
Best Local Similarity 51.7%; Pred. No. 1.7;
Matches 76; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 75 GACGGCAAGCGGAATATGTTTATCCGCAATGAGTGGCGTAAAAACCAATAAAGACAAAT 134
DB 756 GAGGAGGAGAGAAATTTTTCAGACCTCAATTTGATGAAATCCATAGATGACAAAT 815
QY 135 TTATGATGATGTCGGGGAAGATGCCGACCGCAAGACTATGCAAAATATGAAAAACCAAG 194
DB 816 TCTGAGGAAGAAATGGAAGATGCAACCAAAACAAAGCTTTCTAAAAAACAGAGAAAAAG 875
QY 195 TACGGGATCAGGCGATGGATGCAACGAT 221
DB 876 AAACAGAAACGACGACAGAAATATGAT 902

```

## RESULT 4

```

US-09-822-846-188
; Sequence 188, Application US/09822846
; Publication No. US20030027139A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: Lavallie, Edward R.
; APPLICANT: Collins-Racie, Lisa A.
; APPLICANT: Evans, Cheryl
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Agostino, Michael J.
; APPLICANT: Steinger II, Robert J.
; APPLICANT: Bowman, Michael R.
; APPLICANT: Spaulding, Vikki
; APPLICANT: Wong, Gordon G.
; APPLICANT: Clark, Hilary
; APPLICANT: Fechtel, Kim
; APPLICANT: Howes, Steven H.
; APPLICANT: Resnick, Richard J.
; APPLICANT: Gulukota, Kamalakar
; APPLICANT: Graham, James R.
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
; FILE REFERENCE: GIN 5400
; CURRENT APPLICATION NUMBER: US/09/822,846
; CURRENT FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/195,605
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 629
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 188
; LENGTH: 1874
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-822-846-188

```

```

Query Match 14.6%; Score 33.4; DB 11; Length 1874;
Best Local Similarity 51.7%; Pred. No. 1.7;
Matches 76; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 75 GACGGCAAGCGGAATATGTTTATCCGCAATGAGTGGCGTAAAAACCAATAAAGACAAAT 134

```



Db 837 GAGGAGGAGAGAGAAATTTTCAAGACCTCAAAATGATGAAATCCATTAGATGACAA 896  
 QY 135 TTAGATGATCGGGAGAGATGCCGACCGACAGACTATGCAAAATATGAAAAACCAAG 194  
 Db 897 TCTGAGGAGAAATGGAAGATGCACCAAAACAAAGCTTTCTAAAAACAGAGAAAAAG 956  
 QY 195 TACGCGGATCAGCGATGATGACGAT 221  
 Db 957 AAACAGAAACAGACAGAAATATGAT 983

RESULT 5  
 US-10-108-605-348  
 ; Sequence 348, Application US/10108605  
 ; Publication No. US20020160934A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Broadus, Julie  
 ; APPLICANT: Stam, Lynn  
 ; APPLICANT: Bachmann, Jane  
 ; APPLICANT: Kamdar, Kim  
 ; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES FROM DROSOPHILA MELANOGASTER THAT ENCODE  
 ; FILE REFERENCE: 31133B  
 ; CURRENT APPLICATION NUMBER: US/10/108,605  
 ; CURRENT FILING DATE: 2002-03-27  
 ; PRIOR APPLICATION NUMBER: US 09/761,142  
 ; PRIOR FILING DATE: 2001-01-16  
 ; PRIOR APPLICATION NUMBER: US 60/176,418  
 ; PRIOR FILING DATE: 2000-01-14  
 ; NUMBER OF SEQ ID NOS: 361  
 ; SOFTWARE: Patent In Ver. 2.1  
 ; SEQ ID NO 348  
 ; LENGTH: 3891  
 ; TYPE: DNA  
 ; ORGANISM: Drosophila melanogaster  
 US-10-108-605-348

Query Match 14.1%; Score 32.4; DB 13; Length 3891;  
 Best Local Similarity 56.6%; Pred. No. 5;  
 Matches 60; Conservative 0; Mismatches 46; Indels 0; Gaps 0;  
 QY 100 CGCAATGAGTGGCGTAAAAACCAATATAAGACAAATTTAGATGATCGGGGAAGATGCC 159  
 Db 1194 CGAATGACTGGAGGAGCGCGCTCAAGCCATGTCTACCGGAGATGGGAAACTCGCA 1253  
 QY 160 GACCGACAGACTATGCAAAATATGAAAAACCAAGTACGCGGATCA 205  
 Db 1254 CACAGGCAAGCAACTGATACAGCAACACGCGAATTTCTTTCA 1299

RESULT 6  
 US-09-738-973-266  
 ; Sequence 266, Application US/09738973  
 ; Patent No. US20020110563A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Reed, Steven G.  
 ; APPLICANT: Henderson, Robert A.  
 ; APPLICANT: Lodes, Michael J.  
 ; APPLICANT: Fling, Steven P.  
 ; APPLICANT: Mohamath, Raodoh  
 ; APPLICANT: Algate, Paul A.  
 ; APPLICANT: Secrist, Heather  
 ; APPLICANT: Indirias, Carol Yoseph  
 ; APPLICANT: Benson, Darin R.  
 ; APPLICANT: Elliott, Mark  
 ; APPLICANT: Mannion, Jane  
 ; APPLICANT: Kalos, Michael D.  
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
 ; FILE REFERENCE: 210121.475C9  
 ; CURRENT APPLICATION NUMBER: US/09/738,973  
 ; CURRENT FILING DATE: 2000-12-14

; NUMBER OF SEQ ID NOS: 587  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 266  
 ; LENGTH: 552  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-09-738-973-266

Query Match 13.8%; Score 31.6; DB 10; Length 552;  
 Best Local Similarity 47.9%; Pred. No. 4;  
 Matches 91; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 36 AACCTGATTACGGGTGATGACGACGACTTGAACATTTCGCGACGGCAAGCGGAATATGTT 95  
 Db 95 AAGAAGCTGGCGCGCGACGACGACACTTCTCAGTCGCGCGCTGCAGTTCACAGAAGAA 154  
 QY 96 TATCCGCAATGATGGCGTAAAAACCAATTAAGACAAATTTAGATGATGTCGGGGAAGAT 155  
 Db 155 AAGCTTTGGCCAGGCTGAGAAGACAGAAATTTGGATGCTCACTTAGAGAACTCCTTAGCAAA 214  
 QY 156 GCCCGACCGACAGACTATGCAAAATATGAAAAACCAAGTACGCGATCAGGCATGGATG 215  
 Db 215 GCTGATGTACCAAAATATGGACAGAAAAAATAATGAACAAACTGAAGTGTATTGCGAG 274  
 QY 216 CACGATCCAA 225  
 Db 275 CCAATCCAA 284

RESULT 7  
 US-09-854-133-266  
 ; Sequence 266, Application US/09854133  
 ; Publication No. US20020183499A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lodes, Michael J.  
 ; APPLICANT: Mohamath, Raodoh  
 ; APPLICANT: Henderson, Robert A.  
 ; APPLICANT: Benson, Darin R.  
 ; APPLICANT: Secrist, Heather  
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
 ; FILE REFERENCE: 210121.475C10  
 ; CURRENT APPLICATION NUMBER: US/09/854,133  
 ; CURRENT FILING DATE: 2001-05-11  
 ; NUMBER OF SEQ ID NOS: 735  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 266  
 ; LENGTH: 552  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-09-854-133-266

Query Match 13.8%; Score 31.6; DB 10; Length 552;  
 Best Local Similarity 47.9%; Pred. No. 4;  
 Matches 91; Conservative 0; Mismatches 99; Indels 0; Gaps 0;  
 QY 36 AACCTGATTACGGGTGATGACGACGACTTGAACATTTCGCGACGGCAAGCGGAATATGTT 95  
 Db 95 AAGAAGCTGGCGCGCGACGACGACACTTCTCAGTCGCGCGCTGCAGTTCACAGAAGAA 154  
 QY 96 TATCCGCAATGATGGCGTAAAAACCAATTAAGACAAATTTAGATGATGTCGGGGAAGAT 155  
 Db 155 AAGCTTTGGCCAGGCTGAGAAGACAGAAATTTGGATGCTCACTTAGAGAACTCCTTAGCAAA 214  
 QY 156 GCCCGACCGACAGACTATGCAAAATATGAAAAACCAAGTACGCGATCAGGCATGGATG 215  
 Db 215 GCTGAATGTACCAAAATATGGACAGAAAAAATAATGAACAAACTGAAGTGTATTGCGAG 274  
 QY 216 CACGATCCAA 225  
 Db 275 CCAATCCAA 284

```
RESULT 8
US-10-144-649A-266
; Sequence 266, Application US/10144649A
; Publication No. US20030118599A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Wang, Tongtong
; APPLICANT: Fan, Liqun
; APPLICANT: Algate, Paul A.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.475C11
; CURRENT APPLICATION NUMBER: US/10/144.649A
; CURRENT FILING DATE: 2002-08-21
; NUMBER OF SEQ ID NOS: 749
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 266
; LENGTH: 552
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-144-649A-266

Query Match 13.8%; Score 31.6; DB 14; Length 552;
Best Local Similarity 47.9%; Pred. No. 4;
Matches 91; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 36 AACCTGATTACGGTGATGACGAGCAGCTTGAACATTCGCGACGCAAGCGGGAATGTT 95
DB 95 AAGAGCTGGCGCGGACGAGGACCTTCTCAGTCGCGCGTGCGAGTTCACAGAGAA 154
QY 96 TATCCGCAATGAGTGGCGTAAACCAATAAAGACAAATTTAGATGATGTCGGGGAAGAT 155
DB 155 AAGCTTGCCAGGCTGAGACACAGAATTTGGATGCTCCTCAGTCGCGCGTGCGAGTTCACAGAGAA 214
QY 156 GCCGACCGCAAGACTATGAAATATGAAACCAATTAAGACAAATTTAGATGATGTCGGGGAAGAT 215
DB 215 GCTGAATGTACCAATATGACAGAGAAAATAATGAACAACTGAAGTGTATTGCAG 274
QY 216 CACGATCCAA 225
DB 275 CCAATCCAA 284

RESULT 9
US-10-044-090-43
; Sequence 43, Application US/10044090
; Publication No. US20020137081A1
; GENERAL INFORMATION:
; APPLICANT: Olga Bandman
; TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
; FILE REFERENCE: PA-0028 US
; CURRENT APPLICATION NUMBER: US/10/044.090
; CURRENT FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: PERL Program
; SEQ ID NO 43
; LENGTH: 1538
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20020137081A1 1990126CB1
US-10-044-090-43

Query Match 13.8%; Score 31.6; DB 13; Length 1538;
Best Local Similarity 47.9%; Pred. No. 6.2;
Matches 91; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 36 AACCTGATTACGGTGATGACGAGCAGCTTGAACATTCGCGACGCAAGCGGGAATGTT 95
DB 99 AAGAGCTGGCGCGGACGAGGACCTTCTCAGTCGCGCGTGCGAGTTCACAGAGAA 158

RESULT 10
US-09-738-973-224
; Sequence 224, Application US/09738973
; Patent No. US20020110563A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Fling, Steven P.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Algate, Paul A.
; APPLICANT: Secrist, Heather
; APPLICANT: Indrias, Carol Yoseph
; APPLICANT: Benson, Darin R.
; APPLICANT: Elliot, Mark
; APPLICANT: Mannion, Jane
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.475C9
; CURRENT APPLICATION NUMBER: US/09/738.973
; CURRENT FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 587
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 224
; LENGTH: 1583
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-738-973-224

Query Match 13.8%; Score 31.6; DB 10; Length 1583;
Best Local Similarity 47.9%; Pred. No. 6.2;
Matches 91; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 36 AACCTGATTACGGTGATGACGAGCAGCTTGAACATTCGCGACGCAAGCGGGAATGTT 95
DB 131 AAGAAGCTGGCGCGGACGAGGACCTTCTCAGTCGCGCGTGCGAGTTCACAGAGAA 190
QY 96 TATCCGCAATGAGTGGCGTAAACCAATAAAGACAAATTTAGATGATGTCGGGGAAGAT 155
DB 191 AAGCTTGCCAGGCTGAGAGACAGAATTTGGATGCTCCTTAGAGAACTCTTAGCAAA 250
QY 156 GCCGACCGCAAGACTATGAAAATATGAAAAACCAAGTACGCGGATCAGGCAATGATG 215
DB 251 GCTGAATGTACCAATATGACAGAGAAAATAATGAACAACTGAAGTGTATTGCAG 310
QY 216 CACGATCCAA 225
DB 311 CCAATCCAA 320

RESULT 11
US-09-854-133-224
; Sequence 224, Application US/09854133
; Publication No. US20020183499A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Henderson, Robert A.
; APPLICANT: Benson, Darin R.
```

```

; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; FILE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.475C10
; CURRENT APPLICATION NUMBER: US/09/854.133
; CURRENT FILING DATE: 2001-05-11
; NUMBER OF SEQ ID NOS: 735
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 224
; LENGTH: 1583
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-854-133-224

Query Match      13.8%; Score 31.6; DB 10; Length 1583;
Best Local Similarity 47.9%; Pred. No. 6.2;
Matches 91; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 36 AACCTGATTACGGGTGATCGAGCGAGCTTGAAACATTCGCGACGGCAAGCGGGAATATGTT 95
Db 131 AAGAGCTGGCGCGCGAGCGAGCACCTTCCTCAGTCGCGCGTGCAGTTCACAGAAGAA 190

QY 96 TATCCGCAATGAGTGGCGTAAACCAATAAAGACAAATTTAGATGATGTCGGGGAAGAT 155
Db 191 AAGCTTGCCCGAGGCTGAGAGACAGATTTGGATGCTCACTTAGAGAACTCCTTAGCAAA 250

QY 156 GCCGACCGACAGAGACTATGCAAAATATGAAAACCAAGTACGGCGGATCGGCATGGATG 215
Db 251 CTTGAATGTACAAATATGGACAGAAAATAATGAACAACTGAAGTGTATTGCAG 310

QY 216 CACGATCCAA 225
Db 311 CCAAAATCCAA 320

RESULT 12
US-10-144-649A-224
; Sequence 224, Application US/10144649A
; Publication No. US20030118599A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Wang, Tongtong
; APPLICANT: Fan, Liqun
; APPLICANT: Algate, Paul A.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; FILE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.475C11
; CURRENT APPLICATION NUMBER: US/10/144.649A
; CURRENT FILING DATE: 2002-08-21
; NUMBER OF SEQ ID NOS: 749
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 224
; LENGTH: 1583
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-144-649A-224

Query Match      13.8%; Score 31.6; DB 14; Length 1583;
Best Local Similarity 47.9%; Pred. No. 6.2;
Matches 91; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 36 AACCTGATTACGGGTGATCGAGCGAGCTTGAAACATTCGCGACGGCAAGCGGGAATATGTT 95
Db 131 AAGAGCTGGCGCGCGAGCGAGCACCTTCCTCAGTCGCGCGTGCAGTTCACAGAAGAA 190

QY 96 TATCCGCAATGAGTGGCGTAAACCAATAAAGACAAATTTAGATGATGTCGGGGAAGAT 155
Db 191 AAGCTTGCCCGAGGCTGAGAGACAGATTTGATGCTCACTTAGAGAACCTCCTTAGCAAA 250

QY 156 GCCGACCGACAGACTATGCAAAATATGAAAACCAAGTACGGCGGATCGGCATGGATG 215
Db 251 CTTGAATGTACAAATATGGACAGAAAATAATGAACAACTGAAGTGTATTGCAG 310

QY 216 CACGATCCAA 225
Db 311 CCAAAATCCAA 320

RESULT 13
US-09-764-891-9126
; Sequence 9126, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: FC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9126
; LENGTH: 12100
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-9126

Query Match      13.7%; Score 31.4; DB 11; Length 12100;
Best Local Similarity 53.7%; Pred. No. 17;
Matches 65; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY 104 ATGAGTGGCGTAAAAACCAATTAAGACAAATTTAGATGATGTCGGGGAAGATGCCCGACC 163
Db 4243 AAGATGAGAGAGAGATTCAAATTAAGAAAAATTAGAATGATGAAGGGAATGTTACCACTGA 4302

QY 164 GACAAGACTATGCAAAATATGAAAACCAAGTACGGCGGATCGGCATGGATGACGATCC 223
Db 4303 CCCAGAGAAAAAAAATAACCAACCAAGTAACTACGAGCACCTCTATACACACAAACT 4362

QY 224 A 224
Db 4363 A 4363

RESULT 14
US-10-091-572-702
; Sequence 702, Application US/10091572
; Publication No. US20030054373A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA118C1
; CURRENT APPLICATION NUMBER: US/10/091,572
; CURRENT FILING DATE: 2002-03-07
; Prior application number: 09/764,850
; Prior filing date: 2001-01-17
; Prior application number: 60/179,065
; Prior filing date: 2000-01-31
; Prior application number: 60/180,628
; Prior filing date: 2000-02-04
; Prior application number: 60/214,886
; Prior filing date: 2000-06-28
; Prior application number: 60/217,487
; Prior filing date: 2000-07-11
; Prior application number: 60/225,758
; Prior filing date: 2000-08-14
; Prior application number: 60/220,963
; Prior filing date: 2000-07-26
; Prior application number: 60/217,496
; Prior filing date: 2000-07-11
; Prior application number: 60/225,447
; Prior filing date: 2000-08-14
; Prior application number: 60/218,290
; Prior filing date: 2000-07-14
; Prior application number: 60/225,757
```

PRIOR FILING DATE: 2000-08-14	PRIOR APPLICATION NUMBER: 60/222,868
PRIOR FILING DATE: 2000-08-22	PRIOR APPLICATION NUMBER: 60/226,868
PRIOR FILING DATE: 2000-08-22	PRIOR APPLICATION NUMBER: 60/226,868
PRIOR FILING DATE: 2000-08-22	PRIOR APPLICATION NUMBER: 60/216,647
PRIOR FILING DATE: 2000-07-07	PRIOR APPLICATION NUMBER: 60/225,267
PRIOR FILING DATE: 2000-08-14	PRIOR APPLICATION NUMBER: 60/216,880
PRIOR FILING DATE: 2000-07-07	PRIOR APPLICATION NUMBER: 60/225,270
PRIOR FILING DATE: 2000-08-14	PRIOR APPLICATION NUMBER: 60/251,869
PRIOR FILING DATE: 2000-12-08	PRIOR APPLICATION NUMBER: 60/235,834
PRIOR FILING DATE: 2000-09-27	PRIOR APPLICATION NUMBER: 60/234,274
PRIOR FILING DATE: 2000-09-21	PRIOR APPLICATION NUMBER: 60/234,223
PRIOR FILING DATE: 2000-09-21	PRIOR APPLICATION NUMBER: 60/228,924
PRIOR FILING DATE: 2000-08-30	PRIOR APPLICATION NUMBER: 60/224,518
PRIOR FILING DATE: 2000-08-14	PRIOR APPLICATION NUMBER: 60/236,369
PRIOR FILING DATE: 2000-09-29	PRIOR APPLICATION NUMBER: 60/234,519
PRIOR FILING DATE: 2000-08-14	PRIOR APPLICATION NUMBER: 60/220,964
PRIOR FILING DATE: 2000-07-26	PRIOR APPLICATION NUMBER: 60/241,809
PRIOR FILING DATE: 2000-10-20	PRIOR APPLICATION NUMBER: 60/249,299
PRIOR FILING DATE: 2000-11-17	PRIOR APPLICATION NUMBER: 60/236,327
PRIOR FILING DATE: 2000-09-29	PRIOR APPLICATION NUMBER: 60/241,785
PRIOR FILING DATE: 2000-10-20	PRIOR APPLICATION NUMBER: 60/244,617
PRIOR FILING DATE: 2000-11-01	PRIOR APPLICATION NUMBER: 60/225,268
PRIOR FILING DATE: 2000-08-14	PRIOR APPLICATION NUMBER: 60/236,368
PRIOR FILING DATE: 2000-09-29	PRIOR APPLICATION NUMBER: 60/251,856
PRIOR FILING DATE: 2000-12-08	PRIOR APPLICATION NUMBER: 60/251,868
PRIOR FILING DATE: 2000-12-08	PRIOR APPLICATION NUMBER: 60/229,344
PRIOR FILING DATE: 2000-09-01	PRIOR APPLICATION NUMBER: 60/229,345
PRIOR FILING DATE: 2000-09-01	PRIOR APPLICATION NUMBER: 60/229,287
PRIOR FILING DATE: 2000-09-25	PRIOR APPLICATION NUMBER: 60/229,513
PRIOR FILING DATE: 2000-09-25	PRIOR APPLICATION NUMBER: 60/229,343
PRIOR FILING DATE: 2000-09-01	PRIOR APPLICATION NUMBER: 60/229,345
PRIOR FILING DATE: 2000-09-01	PRIOR APPLICATION NUMBER: 60/229,287
PRIOR FILING DATE: 2000-09-01	PRIOR APPLICATION NUMBER: 60/229,513
PRIOR FILING DATE: 2000-09-05	PRIOR APPLICATION NUMBER: 60/231,413
PRIOR FILING DATE: 2000-09-08	PRIOR APPLICATION NUMBER: 60/229,509
PRIOR FILING DATE: 2000-09-05	PRIOR APPLICATION NUMBER: 60/236,367
PRIOR FILING DATE: 2000-09-29	PRIOR APPLICATION NUMBER: 60/237,039
PRIOR FILING DATE: 2000-10-02	PRIOR APPLICATION NUMBER: 60/237,038
PRIOR FILING DATE: 2000-10-02	PRIOR APPLICATION NUMBER: 60/236,370
PRIOR FILING DATE: 2000-09-29	PRIOR APPLICATION NUMBER: 60/236,802
PRIOR FILING DATE: 2000-10-02	PRIOR APPLICATION NUMBER: 60/236,802

```

; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,244
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/233,064
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/233,063
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,397
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,399
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,401
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/241,808
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,826
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,786
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,221
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/246,475
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/231,243
; PRIOR FILING DATE: 2000-09-08

```

Query Match 13.7%; Score 31.4; DB 14; Length 12100;

Best Local Similarity 53.7%; Pred. No. 17;

Matches 65; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

```

QY 104 ATGAGTGGCGTAATAAACCAATTAACACAAATTTAGATGTCGGGGGAAGATGCCCGACC 163
Db 4243 AAGATGAGAGAGATTCAAATTAACAAATTTAGAAATGATGAAGGGAATGTTACCACTGA 4302
QY 164 CACAAGACATTCGAAAATATGAAAAACCAAGTACCGGATCAGGATCGATCGACGATCC 223
Db 4303 CCCAGAGAAAAAATAATACACAGTAACTACTACGAGCACCTCTATACACAAACT 4362
QY 224 A 224
Db 4363 A 4363

```

# RESULT 15

US-09-764-864-625

; Sequence 625, Application US/09764864

; Patent No. US20020132753A1

; GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

; FILE REFERENCE: PT223

; CURRENT APPLICATION NUMBER: US/09/764,864

; CURRENT FILING DATE: 2001-01-17

; Prior application data removed - consult PALM or file wrapper

; NUMBER OF SEQ ID NOS: 1792

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 625

; LENGTH: 1251

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: SITE

; LOCATION: (1220)

; OTHER INFORMATION: n equals a,t,g, or c

; NAME/KEY: SITE

; LOCATION: (1249)

; OTHER INFORMATION: n equals a,t,g, or c

US-09-764-864-625

Query Match

Best Local Similarity 13.5%; Score 31; DB 10; Length 1251;

Matches 70; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

```

QY 75 GAGGCAAGCGGGAATATGTTTATCCGCAATGAGTGGCGTAATAAACCAATTAACACAAAT 134
Db 808 GAGGAGGAAGAAGAAAATTTTCAAGACCTCAAAATTTGATGAAAATCCATTAGATGACAAT 867
QY 135 TTAGATGATGTCGGGGGAAGATGCCCGACCGACCAAGACTATGCAAAAATATGAAAAACCAAG 194
Db 868 TCTGAGGAAGAATGGAAGATGCACCAAAACAAAAGCTTTCTTAAAAAACAGAAAAAG 927
QY 195 TACCGGATCAGGCA 209
Db 928 AAACAGAAACCAGCA 942

```

Search completed: November 15, 2003, 08:32:02

Job time : 133.778 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 14, 2003, 23:56:24 ; Search time 1562.62 Seconds  
(without alignments)  
9241.609 Million cell updates/sec

Title: US-09-928-457-83  
Perfect score: 353  
Sequence: 1 CGGGAATTCGACGAGATG.....ACGGAGGACCGTTGGCTTT 353

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2887711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.om.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.hum.\*
- 18: em.in.\*
- 19: em.mu.\*
- 20: em.om.\*
- 21: em.or.\*
- 22: em.ov.\*
- 23: em.pat.\*
- 24: em.ph.\*
- 25: em.pl.\*
- 26: em.ro.\*
- 27: em.sts.\*
- 28: em.un.\*
- 29: em.vi.\*
- 30: em.htg.hum.\*
- 31: em.htg.inv.\*
- 32: em.htg.other.\*
- 33: em.htg.mus.\*
- 34: em.htg.pln.\*
- 35: em.htg.rod.\*
- 36: em.htg.mam.\*
- 37: em.htg.vrt.\*
- 38: em.gy.\*
- 39: em.htgo.hum.\*
- 40: em.htgo.mus.\*
- 41: em.htgo.other.\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	353	100.0	353	6	A68912 Sequence 83
2	353	100.0	353	6	BD063012 DNA and s
3	286.8	81.2	852	1	AF169459 Neisseria
4	286.8	81.2	349061	1	AL162753 Neisseria
5	43.6	12.4	1950	10	AF071754 Mus muscu
6	43.4	12.3	102331	1	AJ494981 Escherich
7	43	12.2	7170	1	AF169965 Bacillus
8	39	11.0	139049	9	AC007314 Homo sapi
9	38.8	11.0	328	6	AX301007 Sequence
10	38.6	10.9	521	9	HSAP4H02 Human alpha
11	38.4	10.9	832	6	BD110729 EST and e
12	38.4	10.9	108490	9	AC118480 Homo sapi
13	38.4	10.9	165789	2	AL929389 Danio rer
14	38.2	10.8	164972	2	AC101994 Mus muscu
15	38.2	10.8	342850	1	AP003597 Nostoc sp
16	38	10.8	233330	2	AC112548 Rattus no
17	38	10.8	252353	2	AC094533 Rattus no
18	37.8	10.7	187190	2	AC116412 Mus muscu
19	37.8	10.7	234178	2	AC127449 Rattus no
20	37.6	10.7	81904	9	HS276A23 Homo sapi
21	37.6	10.7	129688	9	HS197B34 Homo sapi
22	37.6	10.7	186239	9	AC112642 Homo sapi
23	37.6	10.7	212399	9	AC084784 Homo sapi
24	37.2	10.5	590	8	AF541915 Arabidops
25	37.2	10.5	126396	8	AC005169 Arabidops
26	37.2	10.5	150620	2	AC055124 Homo sapi
27	37.2	10.5	166905	2	AC025729 Homo sapi
28	37.2	10.5	200638	9	AL356253 Human DNA
29	37	10.5	110000	2	Continuation (2 of
30	37	10.5	138416	9	AC105218 Homo sapi
31	37	10.5	175970	9	AC097488 Homo sapi
32	37	10.5	181845	2	AC026586 Homo sapi
33	37	10.5	190701	9	AC022274 Homo sapi
34	37	10.5	250029	3	AE014820 Plasmodiu
35	36.6	10.4	175352	9	AC069394 Homo sapi
36	36.6	10.4	193709	9	AL330838 Human DNA
37	36.6	10.4	205919	2	AC012279 Homo sapi
38	36.6	10.4	215849	2	AL590654 Mus muscu
39	36.4	10.3	11050	8	AY180106 Zea mays
40	36.4	10.3	23862	2	AY211535 Zea mays
41	36.4	10.3	148775	2	AC109832 Oryza sat
42	36.4	10.3	174438	2	BX248081 Danio rer
43	36.4	10.3	192100	9	AC021810 Homo sapi
44	36.4	10.3	210523	10	AL669945 Mouse DNA
45	36.2	10.3	408	6	AX300942 Sequence

ALIGNMENTS

RESULT 1  
A68912  
LOCUS A68912 353 bp DNA linear PAT 06-MAY-1999  
DEFINITION Sequence 83 from Patent WO9802347.  
ACCESSION A68912  
VERSION A68912.1 GI:4759831  
KEYWORDS  
SOURCE unidentified  
ORGANISM unidentified  
REFERENCE 1 (bases 1 to 353)  
AUTHORS Nassif, X., Tinsley, C., Achtman, M., Ruelle, J., Vinals, C. and  
Merker, P.  
TITLE DNA AND SPECIFIC PROTEINS OR PEPTIDES OF THE NEISSERIA MENINGITIDIS  
SPECIES BACTERIA, METHOD FOR OBTAINING THEM AND THEIR BIOLOGICAL

Pred. No. is the number of results predicted by chance to have a

APPLICATIONS		APPLICATIIONS	
JOURNAL		Patent: WO 9802547-A 83 22-JAN-1998;	
INST NAT SANTE RECH MED (FR)		INST NAT SANTE RECH MED (FR)	
COMMENT		Other publication FR 2751000 19980116.	
FEATURES		Location/Qualifiers	
source		1. .353	
		/organism="unidentified"	
		/mol_type="genomic DNA"	
		/db_xref="taxon:32644"	
BASE COUNT		119 a 82 g 89 t	
ORIGIN			
Query Match		100.0%; Score 353; DB 6; Length 353;	
Best Local Similarity		100.0%; Pred. No. 4.6e-88;	
Matches 353; Conservative		0; Mismatches 0; Indels 0; Gaps 0;	
Qy	1	CGGGAATTCGAGCAGAATGAAGAAAGCAGCGCTTGATAATTTTCATAAAGTTATTGGAAG	60
Db	1	CGGGAATTCGAGCAGAATGAAGAAAGCAGCGCTTGATAATTTTCATAAAGTTATTGGAAG	60
Qy	61	AAAAAGGATTTACCGTCCATTTCCGTTATTCACAATACGGCTGATTACGGAATTCGCCAAA	120
Db	61	AAAAAGGATTTACCGTCCATTTCCGTTATTCACAATACGGCTGATTACGGAATTCGCCAAA	120
Qy	121	GCGTAAAGATTTACGTTAATTGCAACAGATAACCAAGAAAGCTGGAACCAAGTCA	180
Db	121	GCGTAAAGATTTACGTTAATTGCAACAGATAACCAAGAAAGCTGGAACCAAGTCA	180
Qy	181	AGTATTCGGGCAACCGCTTACGGTAGCCGATGTTTGGGAATGGAATGCTTTCCCAA	240
Db	181	AGTATTCGGGCAACCGCTTACGGTAGCCGATGTTTGGGAATGGAATGCTTTCCCAA	240
Qy	241	CATTATTCGAGACACCAAGACGAAACCGATTTTATGCATAGCTGTGCGGAATTTATCTG	300
Db	241	CATTATTCGAGACACCAAGACGAAACCGATTTTATGCATAGCTGTGCGGAATTTATCTG	300
Qy	301	ATATCATTGAACGATTCGCTTGATACCTAAACCGGAGAACCGTTGGCTTT	353
Db	301	ATATCATTGAACGATTCGCTTGATACCTAAACCGGAGAACCGTTGGCTTT	353
RESULT 3		AF169459	
LOCUS		Neisseria meningitidis strain 22491 clone Cml30 unknown sequence.	
DEFINITION		Neisseria meningitidis	
ACCESSION		AF169459	
VERSION		AF169459.1 GI:9754667	
KEYWORDS		Neisseria meningitidis	
SOURCE		Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;	
ORGANISM		Neisseriaceae; Neisseria.	
REFERENCE		1 (bases 1 to 852)	
AUTHORS		Perrin,A., Nassif,X. and Tinsley,C.R.	
TITLE		Identification of regions of the chromosome of Neisseria	
JOURNAL		meningitidis and Neisseria gonorrhoeae which are specific to	
		pathogenic Neisseriae	
		Unpublished	
		2 (bases 1 to 852)	
		Perrin,A., Nassif,X. and Tinsley,C.R.	
		Direct Submission	
		Submitted (16-JUL-1999) Necker Medicine Faculty, INSERM U411, 156	
		rue de Vaugirard, Paris 75015, France	
FEATURES		Location/Qualifiers	
source		1. .852	
		/organism="Neisseria meningitidis"	
		/mol_type="genomic DNA"	
		/strain="22491"	
		/db_xref="taxon:487"	
		/clone="Cml30"	
BASE COUNT		272 a 154 c 206 g 220 t	
ORIGIN			
Query Match		81.2%; Score 286.8; DB 1; Length 852;	
Best Local Similarity		96.6%; Pred. No. 1.6e-69;	
Matches 336; Conservative		0; Mismatches 7; Indels 5; Gaps 4;	

APPLICATIONS		APPLICATIIONS	
JOURNAL		Patent: WO 9802547-A 83 22-JAN-1998;	
INST NAT SANTE RECH MED (FR)		INST NAT SANTE RECH MED (FR)	
COMMENT		Other publication FR 2751000 19980116.	
FEATURES		Location/Qualifiers	
source		1. .353	
		/organism="unidentified"	
		/mol_type="genomic DNA"	
		/db_xref="taxon:32644"	
BASE COUNT		119 a 82 g 89 t	
ORIGIN			
Query Match		100.0%; Score 353; DB 6; Length 353;	
Best Local Similarity		100.0%; Pred. No. 4.6e-88;	
Matches 353; Conservative		0; Mismatches 0; Indels 0; Gaps 0;	
Qy	1	CGGGAATTCGAGCAGAATGAAGAAAGCAGCGCTTGATAATTTTCATAAAGTTATTGGAAG	60
Db	1	CGGGAATTCGAGCAGAATGAAGAAAGCAGCGCTTGATAATTTTCATAAAGTTATTGGAAG	60
Qy	61	AAAAAGGATTTACCGTCCATTTCCGTTATTCACAATACGGCTGATTACGGAATTCGCCAAA	120
Db	61	AAAAAGGATTTACCGTCCATTTCCGTTATTCACAATACGGCTGATTACGGAATTCGCCAAA	120
Qy	121	GCGTAAAGATTTACGTTAATTGCAACAGATAACCAAGAAAGCTGGAACCAAGTCA	180
Db	121	GCGTAAAGATTTACGTTAATTGCAACAGATAACCAAGAAAGCTGGAACCAAGTCA	180
Qy	181	AGTATTCGGGCAACCGCTTACGGTAGCCGATGTTTGGGAATGGAATGCTTTCCCAA	240
Db	181	AGTATTCGGGCAACCGCTTACGGTAGCCGATGTTTGGGAATGGAATGCTTTCCCAA	240
Qy	241	CATTATTCGAGACACCAAGACGAAACCGATTTTATGCATAGCTGTGCGGAATTTATCTG	300
Db	241	CATTATTCGAGACACCAAGACGAAACCGATTTTATGCATAGCTGTGCGGAATTTATCTG	300
Qy	301	ATATCATTGAACGATTCGCTTGATACCTAAACCGGAGAACCGTTGGCTTT	353
Db	301	ATATCATTGAACGATTCGCTTGATACCTAAACCGGAGAACCGTTGGCTTT	353
RESULT 2		BD063012	
LOCUS		353 bp DNA linear PAT 27-AUG-2002	
DEFINITION		DNA and specific proteins or peptides of the Neisseria meningitidis	
		species bacteria, method for obtaining them and their biological	
		applications.	
ACCESSION		BD063012	
VERSION		BD063012.1 GI:22608615	
KEYWORDS		JP 2001504684-A/74.	
SOURCE		unidentified	
ORGANISM		unclassified.	
REFERENCE		1 (bases 1 to 353)	
AUTHORS		Nassif,X., Tinsley,C., Achtman,M., Ruelle,J.L., Vinals,C. and	
		Merker,P.	
TITLE		DNA and specific proteins or peptides of the Neisseria meningitidis	
		species bacteria, method for obtaining them and their biological	
		applications	
JOURNAL		Patent: JP 2001504684-A 74 10-APR-2001;	
		INSTITUT NATIONAL DE LA SANTE ET DE LA RECHERCHE MEDICALE, MAX	
		PLANCK GESELLSCHAFT ZUR FORDERUNG DER WISSENSCHAFTEN EV BERLIN,	
		SMITHKLINE BEECHAM	
COMMENT		PN JP 2001504684-A/74	
		PD 10-APR-2001	
		PR 11-JUL-1997 JP 1998050685	
		PR 12-JUL-1996 FR 96/08768	
		PI XAVIER NASSIF, COLIN TINSLEY, MARK ACHTMAN, JEAN LOUIS RUELLE, PI	
		CARLA VINALS,	
		PI PETRA MERKER	
		PC C12N15/31, C07K14/22, C07K16/12, A61K39/095, C12Q1/68, G01N33/53 CC	
		Strandedness: Single;	
		CC Topology: Linear;	



QY 1 CGGGAATTCGAGCAGATGAAGAAGCAGGCTTGATTAATTCATAAAGTATTGGAAG 60  
Db 1 CGGGAATTCGAGCAGATGAAGAAGCAGGCTTGATTAATTCATAAAGTATTGGAAG 60  
QY 61 AAAAGGATTCACGTCCTCAATTCGGTATTCAAAATACGCTGATTACGGAATTCGCCAAA 120  
Db 61 AAAAGGATTCACGTCCTCAATTCGGTATTCAAAATACGCTGATTACGGAATTCGCCAAA 120  
QY 121 GCGGTAAAGATTTACGTTTAAATGCAACACAGATTAACCAAGAAAGCTGGAACCAAGTCA 180  
Db 121 GCGGTAAAGATTTACGTTTAAATGCAACACAGATTAACCAAGAAAGCTGGAACCAAGTCA 180  
QY 181 AGTATTCGGGCAACCGCTTACGTTAGCGATGTTTGGGAATGG-AAATGCGCTTTCCCA 239  
Db 181 AGTATTCGGGCAACCGCTTACGTTAGCGATGTTTGGGAATGG-AAATGCGCTTTCCCA 240  
QY 240 ACATATTTCAGGACACCAAGCAGCAACCGATTTTATGTCATAGCTGTGCGGAATATCT 299  
Db 241 AAATATTTCAGGACACCAAGCAGCAACCGATTTTATGTCATAGCTGTGCGGG-ATTATCT 299  
QY 300 GATATCA--CITGAACGATTTGCG-TTGATACCTAAACGAGGAAC 344  
Db 300 GATATCACTTGAACGATTTGCGTTGATACCTAAACGAGGAAC 347  
RESULT 4  
NMA222491 349061 bp DNA linear BCT 02-SEP-2002  
LOCUS Neisseria meningitidis serogroup A strain 22491 complete genome;  
DEFINITION segment 2/7.  
ACCESSION AL162753 AL157959  
VERSION AL162753.2 GI:7379120  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
Parkhill,J., Achtman,M., James,K.D., Bentley,S.D., Churcher,C., Klee,S.R., Morelli,G., Basham,D., Brown,D., Chillingworth,T., Davies,R.M., Davis,P., Devlin,K., Feltwell,T., Hamlin,N., Holtroyd,S., Jagels,K., Leather,S., Moule,S., Mungall,K., Quail,M.A., Rajandream,M.A., Rutherford,K.M., Simmonds,M., Skelton,J., Whitehead,S., Spratt,B.G. and Barrell,B.G.  
Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491  
JOURNAL Nature 404 (6777), 502-506 (2000)  
MEDLINE 20222536  
PUBMED 10761919  
REFERENCE 2 (bases 1 to 349061)  
AUTHORS Parkhill,J.  
TITLE Direct Submission  
JOURNAL Submitted (30-MAR-2000) Submitted on behalf of the Neisseria sequencing team, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk  
COMMENT Notes:  
Details of N. meningitidis sequencing at the Sanger Centre are available on the World Wide Web.  
(URL, http://www.sanger.ac.uk/Projects/N\_meningitidis/).  
FEATURES  
source  
1. 349061  
/organism="Neisseria meningitidis 22491"  
/mol\_type="genomic DNA"  
/strain="22491"  
/db\_xref="taxon:122587"  
/note="serogroup: A"  
complement(24..206)  
/note="ATR repeat; hmfs hit to HMM ATR (1 - 183), score: 310.39"  
repeat\_unit  
/label=ATR  
209..212  
RBS  
220..681  
gene

CDS  
/gene="NMA0368"  
220..681  
/note="NMA0368"  
/note="NMA0368, probable integral membrane protein, len: 153 aa; contains four probable transmembrane domains"  
/codon\_start=1  
/transl\_table=11  
/product="putative integral membrane protein"  
/protein\_id="CA883669.1"  
/db\_xref="GI:7379121"  
/db\_xref="SPTREMBL:Q9JWH7"  
/translation="MQEQNRKSSFPVIMLVSVLWIAISNVAFVILGHSGMEGLTVLILGSPASIDIRYCAYANYVWLAIVLILAKRKVPVHAFAFWGLVAFVSKVYVDEAGNTSDIVRYCAGFYLVYAAFAVASIGTFAGKNKKAASAADGTHNDV"  
671..1492  
/gene="hemk"  
671..1492  
/genes="hemk"  
/note="NMA0369, hemk, Hemk protein, len: 273 aa; similar to e.g. HEMK ECOLI P37186 HEMK protein (277 aa), fasta scores: E(): 0, 42.3% identity in 279 aa overlap. Contains PS00092 N-6 Adenine-specific DNA methylases signature"  
/codon\_start=1  
/transl\_table=11  
/product="Hemk protein"  
/protein\_id="CA883670.1"  
/db\_xref="GI:7379122"  
/db\_xref="SPTREMBL:Q9JWH6"  
/translation="MTFDKWLGLSKLPKNEARMLLOYSEYTRVQLITRGSEMPDEVQRADRLAARRLNGCEFAVILGARFYGRFTVNPVSLIPRTEHLVAVLARKLPENGRVDLTGSGAVAVTVALLERDAFVRASDISPPALETKRNAADILGARVEFAYGWFDTMDPSEGKWDIIIVNPPYIENGDKHLSQDLRFEPQIALITDFSDLSGIRTLAQGAPDLAEGGFLLEHGFQGAARVGLAENGFSGVETPLDLAGLDRVTILGKYMHLK"  
831..840  
/gene="hemk"  
/note="Core DNA uptake sequence: gcgcgtctgaa"  
/label=DUS  
1162..1171  
/genes="hemk"  
/note="Core DNA uptake sequence: gcgcgtctgaa"  
/label=DUS  
1187..1207  
/gene="hemk"  
/note="PS00092 N-6 Adenine-specific DNA methylases signature"  
/note="Core DNA uptake sequence: gcgcgtctgaa"  
complement(1279..1288)  
/label=DUS  
1565..1568  
1578..2366  
/gene="NMA0370"  
1578..2366  
/genes="NMA0370"  
/note="NMA0370"  
/note="NMA0370, probable integral membrane protein, len: 462 aa; similar to hypothetical proteins e.g. Y325\_HAEIN P44640 hypothetical protein HI0325 (450 aa), fasta scores: E(): 0, 49.8% identity in 464 aa overlap"  
/codon\_start=1  
/transl\_table=11  
/product="putative integral membrane protein"  
/protein\_id="CA883671.1"  
/db\_xref="GI:7379123"  
/db\_xref="SPTREMBL:Q9JWH5"  
/translation="NNAVVAVIVMLVLSLRVHVLSLTIGAFVGGAVAGPLQNIADAAGQVQAGIIPVFNKLEGAKIALSYAMLGAFAMATHSGLPQQLAGAVVRKLNRGMPPDSVRSEGKVKWLLSILVNMKMSQNIPIHIAFIPVLPVLLVFNRLKIDRRLIACVITTEGLVTFYFLPYGFGAIFLANEILILGNISAAQPDQVKNINWAAAPALGMLALLAFVHYRPRLYQSNNAADTAGNADANRPOPSAYRSLAAVAIAYCFATQLWYEDSLILGAMLGEAFVFMVGLVINEKDANDVFGELKOMAYGVFIWIAOGGAWNNA TGHIOPLVSSMAIFGNSKGMALAMLVGLLVTVMTGIGSSFSTFIILAIYPLCVGLGFSPLATVAIVGTAGALDAGSPASDSTLGTPTMGLNADGQHDIHDSVITPIFIHNIPLLIAGWIAAWV"  
2367..3191  
gene









```

source
1. 139049
/organism="Homo sapiens"
/mol_type="Genomic DNA"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-74G24"
/clone_lib="RPCI-11"
832. 893
/rpt_family="MER1_type"
repeat_region
1376. 1594
/rpt_family="Alu"
repeat_region
2219. 2524
/rpt_family="Alu"
repeat_region
2936. 3239
/rpt_family="Alu"
repeat_region
3286. 3457
/rpt_family="MER1_type"
repeat_region
3475. 4180
/rpt_family="MER21-group"
repeat_region
4228. 4548
/rpt_family="MER1_type"
repeat_region
4543. 4861
/rpt_family="Alu"
repeat_region
4862. 4883
/rpt_family="MER1_type"
repeat_region
4884. 5184
/rpt_family="Alu"
repeat_region
5185. 5564
/rpt_family="MER1_type"
repeat_region
5565. 5875
/rpt_family="Alu"
repeat_region
5876. 6517
/rpt_family="MER1_type"
repeat_region
6518. 6820
/rpt_family="Alu"
repeat_region
6821. 7393
/rpt_family="MER1_type"
repeat_region
7394. 7676
/rpt_family="Alu"
repeat_region
7677. 7697
/rpt_family="(TAAA)n"
repeat_region
7698. 7723
/rpt_family="MER1_type"
repeat_region
7757. 8064
/rpt_family="Alu"
repeat_region
8090. 8597
/rpt_family="MaLR"
repeat_region
8696. 8728
/rpt_family="L1"
repeat_region
8729. 9027
/rpt_family="Alu"
repeat_region
9032. 9079
/rpt_family="(CA)n"
repeat_region
9080. 9206
/rpt_family="L1"
repeat_region
9210. 9352
/rpt_family="Alu"
repeat_region
9353. 9666
/rpt_family="Alu"
repeat_region
9667. 9878
/rpt_family="L1"
repeat_region
9879. 10047
/rpt_family="Alu"
repeat_region
10048. 10126
/rpt_family="(TTTC)n"
repeat_region
10127. 10271
/rpt_family="Alu"
repeat_region
10272. 10299
/rpt_family="L1"
repeat_region
10306. 10595
/rpt_family="Alu"
repeat_region

repeat_region
13524. 13656
/rpt_family="Retroviral"
repeat_region
15083. 15207
/rpt_family="MaLR"
repeat_region
15208. 15496
/rpt_family="MER4-group"
repeat_region
15497. 16118
/rpt_family="Alu"
repeat_region
16185. 16223
/rpt_family="MER4-group"
repeat_region
16437. 16580
/rpt_family="(TC)n"
repeat_region
17545. 17841
/rpt_family="MaLR"
repeat_region
17842. 18143
/rpt_family="Alu"
repeat_region
18399. 18447
/rpt_family="L1"
repeat_region
18499. 18766
/rpt_family="L1"
repeat_region
18769. 19074
/rpt_family="Alu"
repeat_region
20185. 20483
/rpt_family="L1"
repeat_region
20485. 20523
/rpt_family="L2"
repeat_region
20935. 21173
/rpt_family="MaLR"
repeat_region
21185. 21340
/rpt_family="L2"
repeat_region
21341. 21644
/rpt_family="Alu"
repeat_region
21645. 21667
/rpt_family="L2"
repeat_region
22413. 22475
/rpt_family="L2"
repeat_region
23183. 23295
/rpt_family="MIR"
repeat_region
24085. 24140
/rpt_family="MIR"
repeat_region
24589. 24734
/rpt_family="MER1_type"
repeat_region
24737. 24845
/rpt_family="L2"
repeat_region
26770. 27073
/rpt_family="Alu"
repeat_region
27395. 27631

Query Match 11.0%; Score 39; DB 9; Length 139049;
Best Local Similarity 48.8%; Pred No. 3.7;
Matches 105; Conservative 0; Mismatches 110; Indels 0; Gaps 0;

QY 14 CAGAATGAAGAAGCAGCGTTGATAATTTTCAATAAGTTATTGGAAGAAAAGGATTTAC 73
Db 47536 CAGTTTGAGAGATACCAGCGTTATTGTGTGTCAGATCAGCAATGGAAGATTTAG 47477
QY 74 CGTCCATTTCGGTATTCACATACGGCTGATTACGGATTCCTCCCAAGCCGTAAGATT 133
Db 47476 CGTCATTGTATAAGTAAGATTAGGCTAAGATGGAAGCTTGGGAAAAACAATATTT 47417
QY 134 TACGTTAATTGCAACAGATAAACCAAGAAAAGTGGAACCAAGTCAGTCAAGTATTCGGCAA 193
Db 47416 TACTTGAATCTGTGATAAAATCCATACACCTCTCTAGAGAAAATACAAATGCT 47357
QY 194 ACGGTTACGGTAGCCGATGTTTGGGAATGAAA 228
Db 47356 GTGATGTATTTACATAGATTTTAGAGAAATTCACA 47322

RESULT 9
AX301007/c AX301007 328 bp DNA linear PAT 30-NOV-2001
LOCUS

```

DEFINITION Sequence 78 from Patent WO0184903.  
 AX301007  
 VERSION AX301007.1 GI:17382272  
 KEYWORDS  
 SOURCE synthetic construct  
 ORGANISM synthetic construct  
 artificial sequences.

REFERENCE 1  
 AUTHORS Donne-Gousse,C., Laudet,V. and Hanni,C.  
 TITLE Method for detecting and identifying the presence of biological  
 JOURNAL substances derived from birds, and oligonucleotides therefor  
 Patent: WO 0184903-A 78 15-NOV-2001;  
 CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE (CNRS) (FR)

FEATURES  
 Location/Qualifiers  
 1..328  
 /organism="synthetic construct"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:32630"  
 /note="fragment d'ADN amplifie partir du genome d'oiseaux"

BASE COUNT 34 a 22 c 14 g 21 t 237 others

Query Match 11.0%; Score 38.8; DB 6; Length 328;  
 Best Local Similarity 11.9%; Pred. No. 4.7;  
 Matches 36; Conservative 149; Mismatches 117; Indels 0; Gaps 0;

QY 7 TTCTGAGCAGATGAAGAACGACGCTTGATATTCATTAAGTTATGGAAGAAAAG 66  
 Db 328 DNDKKRWSWAGGRYRDDDBDKYVYRRYDRDHYDRBNVDKTKNKRNRDB 269

QY 67 GATTACCGTCGATTCGATTCACAAATACGGCTGATTACGGAATCCCAAGCCGTA 126  
 Db 268 DRRDTRTRKDDWDYDNRDXTDNDWBHDBBYTRGTRCTGGMGTTCTTAAATAADKW 209

QY 127 AAAGATTACGTTAATTCGCAACAGATAACCAAGAAAGCTGGAACCACTCAAGTATT 186  
 Db 208 TRDDHHRHHRHTRHWDYKHWYNDGWHKHVRVNDVBNVNRDWRSRKRWTDNRN 149

QY 187 CGGCAAAACGGCTACCGTAGCGGATGTTTGGGAATGGAATGCCATTCACCAATAT 246  
 Db 148 DGVRTDWDYRNDNRDNDNDDBDDKDDYDGTGTHKRTGTHKRTYKNDKDTATG 89

QY 247 TCGAGCACCAAGAACGAGATTATTCATGATAGCTGTGGGGAATATCTCATATCA 306  
 Db 88 TCCDRCAGCATTCACYNRRYARCAYBWDVNVNTRGRDBDGYGGANRRRCWYAACYRA 29

QY 307 CT 308  
 Db 28 AT 27

RESULT 10  
 HSAP4H02/c  
 LOCUS HSAP4H02 521 bp DNA linear PRI 07-JAN-1995  
 DEFINITION Human alpha-subunit of prollyl 4-hydroxylase gene, exon 2.  
 ACCESSION U14607  
 VERSION U14607.1 GI:602659  
 KEYWORDS  
 SEGMENT  
 SOURCE 2 of 16  
 Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 521)  
 AUTHORS HeLaakoski,T., Veijola,J., Vuori,K., Rehn,M., Chow,L.T.,  
 TITLE Tailon-Miller,P., Kivirikko,K.I. and Pihlajaniemi,T.  
 Structure and expression of the human gene for the alpha subunit of  
 prollyl 4-hydroxylase. The two alternatively spliced types of mRNA  
 correspond to two homologous exons the sequences of which are  
 expressed in a variety of tissues  
 J. Biol. Chem. 269 (45), 27847-27854 (1994)  
 JOURNAL 95050550  
 MEDLINE  
 PUBMED 7961714

REFERENCE 2 (bases 1 to 521)  
 AUTHORS Rehn,M.  
 TITLE Direct Submission  
 JOURNAL Submitted (12-SEP-1994) Marko Rehn, Department of Medical  
 Biochemistry, University of Oulu, Kajaanintie 52 A, Oulu, Finland,  
 Fin-90220

FEATURES  
 Location/Qualifiers  
 1..521  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 <1..208  
 /number=1  
 /cons\_splices=(5'site:NO, 3'site:YES)  
 /evidence=experimental  
 209..316  
 /product="alpha-subunit of prollyl 4-hydroxylase"  
 /number=2  
 /evidence=experimental  
 317..521  
 /number=2  
 /cons\_splices=(5'site:YES, 3'site:NO)  
 /evidence=experimental

BASE COUNT 161 a 76 c 213 t

ORIGIN

Query Match 10.9%; Score 38.6; DB 9; Length 521;  
 Best Local Similarity 46.5%; Pred. No. 5.3;  
 Matches 125; Conservative 0; Mismatches 144; Indels 0; Gaps 0;

QY 7 TTCTGACGAGATGAAGAACGACGCTTGATATTCATTAAGTTATGGAAGAAAAG 66  
 Db 490 TGTATAATAAACGTAACGAAATGTCAAATATTATAAAAGGACCTTTAGGGCAGG 431

QY 67 GATTACCGTCCATTTCCGTTATTCACAAATACGGCTGATTACGGAATCCCAAGCCGTA 126  
 Db 430 GAATAAACCTATCTACTCGAATACACTGATTGTTGTTGCTGATTTAAACACTAAA 371

QY 127 AAAGATTACGTTAATTCGCAACAGATAACCAAGAAAGCTGGAACCACTCAAGTATT 186  
 Db 370 AATGGAATTAATAAATGCAAAACCAACCAACCAACCAAGTAGTAGTACCAATTG 311

QY 187 CGGCAAAACGGCTACCGTAGCGGATGTTTGGGAATGGAATGGTTTCCCAACATTAT 246  
 Db 310 AAGTAAAAAGCCCTGGATGAGCAAGACTGGGGAAGCAGAAATTCCTATAATTAATAT 251

QY 247 TCGAGGACACCAAGACGAAACGGATTTA 275  
 Db 250 ACCAGATCATCTTGGGAAGATTAAATTTA 222

RESULT 11  
 BD110729/c  
 LOCUS BD110729 832 bp DNA linear PAT 18-SEP-2002  
 DEFINITION EST and encoded human protein.  
 ACCESSION BD110729  
 VERSION BD110729.1 GI:23205547  
 KEYWORDS JP 2002010789-A/2806.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 832)  
 AUTHORS Edwards,J.B.D.M., Jobert,S. and Giordano,J.E.  
 TITLE EST and encoded human protein  
 JOURNAL Patent: JP 2002010789-A 2806 15-JAN-2002;  
 GENSET CORP  
 COMMENT OS Homo sapiens (human)  
 PN JP 2002010789-A/2806  
 PD 15-JAN-2002  
 PF 07-AUG-2000 JP 2000280989  
 PR 05-AUG-1999 US 60/147499  
 PI JEAN BAPTISTE DUMAS MILNE EDWARDS, SEVELIN JOBERT, JEAN EVE PI

```
GIORDANO
PC C12N15/09,C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,PC
C12N1/21,
PC C12N5/10,C12P21/02,C12P21/08,C12Q1/68,C12N15/00,C12N5/00,PC
C12N15/00
CC EST and encoded human protein
FH Key Location/Qualifiers
FT CDS 235..399.

FEATURES
    source
        1..832
        /organism="Homo sapiens"
        /mol_type="genomic DNA"
        /db_xref="taxon:9606"
BASE COUNT 164 a 114 g 186 t 271 others
ORIGIN
    Query Match 10.9%; Score 38.4; DB 6; Length 832;
    Best Local Similarity 10.3%; Pred No. 6;
    Matches 30; Conservative 138; Mismatches 124; Indels 0; Gaps 0;
QY 13 GCAGATGAAGAAGCAGGCTTGATATTTTCATAAGTTATTGGAAGAAAAAGGATTTA 72
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 320 RYKXYTWGWWYMKRMSTMWYCMWCKCMYGRRCAMWYTWARGWWSYANGKWSMR 261
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 73 CGGTCATTTCGGTATTCACATACGCTGATTACGGAATCCCAAGCCGTAAAGAT 132
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 260 SAMSNCTRMYYKSGTYMTWKTCTATCYWYKWKWMSKTCWSGSGRGFTSYTSTRS 201
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 133 TTACGTTAATGCACAGATAACCAAGAAAGCTGGAACACAGTCAAGTATTCGGCA 192
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 200 YSMYASWMTYCWGWGGRWSTWYMWGKKWRYATTWRAMWMAWMTWYMWMAW 141
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 193 AACGGCTACGGTACCGATGTTTGGGAATGGAATGGCTTCCCAACATTATTCGAGG 252
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 140 CMSRGAANYRTMMWGYRYWRKKSRYRTCAWAYAWKTKRSYWCWKKWRCMMMM 81
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 253 ACACAGACGAAACGGAATTTATGCTAGCTGCGGGAAATTATCTGATAT 304
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 80 MYAAYGKTMMWRACWKTRYWEWMAWMAWMTWMTWMMYMYWYWKRMW 29
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 12
AC118480
LOCUS Homo sapiens BAC clone RP11-393K13 from 4, complete sequence.
AC118480 AC068518
ACCESSION AC118480.3 GI:20279511
VERSION HTG.
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 108490)
AUTHORS Sulston,J.E. and Waterston,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
PUBMED 9847074
REFERENCE
2 (bases 1 to 108490)
AUTHORS Tomlinson,C. and Kozlowski,A.
TITLE The sequence of Homo sapiens BAC clone RP11-393K13
JOURNAL Unpublished (2001)
REFERENCE
3 (bases 1 to 108490)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (17-APR-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 108490)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (18-APR-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
```

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

COMMENT

MO 63108, USA  
5 (bases 1 to 108490)  
Waterston,R.H.  
Direct Submission  
Submitted (24-APR-2002) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
6 (bases 1 to 108490)  
Waterston,R.  
Direct Submission  
Submitted (24-MAY-2002) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Apr 24, 2002 this sequence version replaced gi:20177784.  
----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc>  
Contact: [sapiens@wustl.edu](mailto:sapiens@wustl.edu)  
----- Summary Statistics  
Center project name: H\_NH0393K13  
Drafting Center: WIBR  
-----

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

## MAPPING INFORMATION:

The mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

## SOURCE INFORMATION:

The RP11-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>  
VECTOR: pBACe3.6

## NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-208N20, 2000 bp overlap; the clone sequenced to the right is RP11-286D5, 2000 bp overlap. Actual start of this clone is at base position 112279 of RP11-208N20; actual end is at base position 37031 of RP11-286D5.

Polymorphisms have been identified between AC096774 and AC118480.

The sequence of AC068518 has been incorporated into AC118480.

Location/Qualifiers  
1..108490  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="4"  
/map="4"  
/clone="RP11-393K13"  
/clone\_lib="RP11-11"

repeat\_region  
1..476  
/rpt\_family="L1"  
repeat\_region  
477..1107



```
repeat_region      /rpt_family="ERV1"
1108. .2507
repeat_region      /rpt_family="L1"
2504. .3813
repeat_region      /rpt_family="L1"
3812. .6922
repeat_region      /rpt_family="L1"
7432. .7608
repeat_region      /rpt_family="L1"
8304. .8343
repeat_region      /rpt_family="MIR"
8574. .8778
repeat_region      /rpt_family="MaLR"
8792. .9128
repeat_region      /rpt_family="MaLR"
9150. .9259
repeat_region      /rpt_family="MaLR"
9864. .9914
repeat_region      /rpt_family="(CA)n"
10266. .10293
repeat_region      /rpt_family="AT-rich"
10328. .10713
repeat_region      /rpt_family="L1"
10722. .11327
repeat_region      /rpt_family="MaLR"
11558. .11733
repeat_region      /rpt_family="MaLR"
12039. .12136
repeat_region      /rpt_family="L2"
12888. .12914
repeat_region      /rpt_family="AT-rich"
13161. .13225
repeat_region      /rpt_family="(TA)n"
13335. .13417
repeat_region      /rpt_family="L2"
13592. .13679
repeat_region      /rpt_family="L2"
14297. .14511
repeat_region      /rpt_family="L1"
14611. .14651
repeat_region      /rpt_family="AT-rich"
14680. .15364
repeat_region      /rpt_family="L1"
15365. .15663
repeat_region      /rpt_family="Alu"
15664. .16059
repeat_region      /rpt_family="L1"
16088. .16434
repeat_region      /rpt_family="L1"
16483. .17499
repeat_region      /rpt_family="L1"
17500. .17605
repeat_region      /rpt_family="(TA)n"
17806. .17972
repeat_region      /rpt_family="L1"
17980. .18122
repeat_region      /rpt_family="L1"
18123. .18145
repeat_region      /rpt_family="(TA)n"
18166. .18171
repeat_region      /rpt_family="L1"
18172. .18597
repeat_region      /rpt_family="MaLR"
18598. .20269
repeat_region      /rpt_family="L1"
20265. .20713
repeat_region      /rpt_family="L1"
20765. .20805
repeat_region      /rpt_family="L1"
20806. .21095
repeat_region      /rpt_family="Alu"
21096. .21194
repeat_region      /rpt_family="L1"

repeat_region      21193. .21879
/rpt_family="L1"
21892. .22316
/rpt_family="L1"
22317. .22621
/rpt_family="Alu"
22622. .23339
/rpt_family="L1"
23357. .24625
/rpt_family="ERV1"
24634. .24814
/rpt_family="L1"
24854. .25176
/rpt_family="ERV1"
25275. .25622
/rpt_family="L1"
25625. .25649
/rpt_family="(TA)n"
25783. .25898
/rpt_family="L1"
25899. .25986
/rpt_family="ERV1"
25991. .26402
/rpt_family="L1"

Query Match      10.9%; Score 38.4; DB 9; Length 108490;
Best Local Similarity 51.1%; Pred. No. 5.4;
Matches 90; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

QY 10 TGAGCAGAATGAAGAAGCAGGCTTGATATTTTCATAAGTTATTGGAGAAAAAGGAT 69
|||||
Db 15049 TGAGGAGAACAAAAGAAATAAGAAATTAATCAGACATAAATGAAATATTAACATAAGAT 15108.

QY 70 TTACCGTCATTTCGGTATTTCACATACAGGCTGATTACGGAATCCCCAAGCGTAAAA 129
|||||
Db 15109 TAAAAAATGAATAGAGCATCAACAAAATCAAAAGGTTGGTTTATGAAGATAGTAATAA 15168

QY 130 GATTTCAGTTAATTGCCAAACAGAAATAACCAAGAAAAAGCTGGAACCAAGTCAAGTAT 185
|||||
Db 15169 TTCTTAATCTGTAGCAAGCCCGAGCAAGAAAAAGACATGAACCATTAATCAAT 15224

RESULT 13
AL929389/c
LOCUS
DEFINITION
  Danio rerio clone CH211-51H4, *** SEQUENCING IN PROGRESS ***, 3
  unordered pieces.
ACCESSION
  AL929389
VERSION
  AL929389.4 GI:30025397
KEYWORDS
  HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE
  Danio rerio (zebrafish)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
  Cypriniformes; Cyprinidae; Danio.
  Andrew, R.
REFERENCE
  1 (bases 1 to 165789)
AUTHORS
  Direct Submission
  Submitted (22-MAY-2003) Wellcome Trust Sanger Institute, Hinxton,
  Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
  zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
  On Apr 19, 2003 this sequence version replaced gi:28268589.
  ----- Genome Center
  Center: Wellcome Trust Sanger Institute
  Web site: http://www.sanger.ac.uk
  Contact: zfish-help@sanger.ac.uk
  ----- Project Information
  Center project name: zCSIH4
  ----- Summary Statistics
  Assembly program: XGAP4; version 4.5
  Chemistry: Dye-terminator; 100% of reads
  Consensus quality: 155008 bases at least Q40
```

Consensus quality: 165145 bases at least Q30  
 Consensus quality: 165247 bases at least Q20  
 Insert size: 165589; sum-of-contigs  
 Insert size: 160385; 3.5% error; agarose-fp  
 Quality coverage: 8.68x in Q20 bases; sum-of-contigs  
 coverage: 8.96x in Q20 bases; agarose-fp

\*\*\*\*\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 3 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence.  
 \* as soon as it is available and the accession number will  
 \* be preserved.

\*\*\*\*\*  
 \* 1 69969: contig of 69969 bp in length  
 \* 69970 70069: gap of 100 bp  
 \* 70070 92004: contig of 21935 bp in length  
 \* 92005 92104: gap of 100 bp  
 \* 92105 165789: contig of 73685 bp in length.

## FEATURES

Location/Qualifiers  
 1..165789  
 /organism="Danio rerio"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:7955"  
 /clone="CH211-51H4"  
 /clone\_lib="CHORI-211"  
 misc\_feature  
 1..69969  
 /note="assembly fragment:01530  
 fragment chain:1"  
 70070..92004  
 /note="assembly fragment:00845  
 fragment chain:1"  
 misc\_feature  
 92105..165789  
 /note="assembly fragment:00602"  
 50716 a 30908 c 30918 g 53047 t 200 others

BASE COUNT 50716 a 30908 c 30918 g 53047 t 200 others  
 ORIGIN

Query Match 10.9%; Score 38.4; DB 2; Length 165789;  
 Best local Similarity 51.1%; Pred. No. 5.4;  
 Matches 90; Conservative 0; Mismatches 86; Indels 0; Gaps 0;  
 QY 1 CGGAATTCGAGCAGATGAAGAAAGCAGGCTGATATTCATAAAGTTATTGGAAG 60  
 Db 134880 CGACAAGTTTCAGTAATAAATAAATAAATCTGTAAATTCACAAAATAAACA 134821  
 QY 61 AAAAGAGTTTACGTCATTCGCTATTCACATACGCTGATTACGGGATTCGCCA 120  
 Db 134820 AAAAACAATCTACTTTCGCTAAGATTTTATTCGAATATGATTCCTCAGTATATAAT 134761  
 QY 121 GCCGTAAAGATTTACGTTTATTCGAACAGATTAACCAAGAAAGCTGGAACCA 176  
 Db 134760 ACTGTAATACACTGATGATTCATCCACAGCTAATCTAAGATCAGCTGCTTGCA 134705

RESULT 14  
 AC101994  
 LOCUS Mus musculus clone RP24-406F14, WORKING DRAFT SEQUENCE, 4 unordered  
 DEFINITION pieces.  
 AC101994  
 ACCESSION AC101994  
 VERSION AC101994.2 GI:22381367  
 KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 164972)  
 Birren,B., Nusbaum,C. and Lander,E.  
 Mus musculus, clone RP24-406F14  
 Unpublished  
 TITLE  
 REFERENCE 2 (bases 1 to 164972)  
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,

Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B.,  
 Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,  
 Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,  
 Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,  
 Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,  
 Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
 Hagos,B., Heatford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,  
 Jones,C., Kamat,A., Karatas,A., Kells,C., LaRoque,K.,  
 Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,  
 MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,  
 McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrum,J.,  
 Meneu,L., Mihova,T., Menga,V., Murphy,T., Naylor,J., Nguyen,C.,  
 Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,  
 Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,  
 Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,  
 Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,  
 Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
 Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
 Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,  
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,M.J., Young,G.,  
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
 Direct Submission  
 Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 164972)  
 Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,  
 Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,  
 Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,  
 Cooke,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
 Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,  
 Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,  
 Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,  
 Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,  
 Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,  
 McCarthy,M., Meldrum,J., Meneu,L., Mihova,T., Menga,V.,  
 Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,  
 O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,  
 Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,  
 Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,  
 Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,  
 Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,  
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,  
 Zembek,L., Zimmer,A. and Zody,M.  
 Direct Submission  
 Submitted (21-AUG-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Aug 21, 2002 this sequence version replaced gi:17060770.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: L17843  
 Center clone name: 406.F.14  
 ----- Summary Statistics  
 Sequencing vector: plasmid; n/a; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 164289 bases at least Q40  
 Consensus quality: 164529 bases at least Q30  
 Consensus quality: 164587 bases at least Q20  
 Insert size: 151000; agarose-fp  
 Insert size: 164672; sum-of-contigs  
 Quality coverage: 9.7 in Q20 bases; agarose-fp  
 Quality coverage: 8.9 in Q20 bases; sum-of-contigs

\*\*\*\*\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 4 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 50: contig of 50 bp in length  
 \* 150: gap of 100 bp  
 \* 151 27038: contig of 26888 bp in length  
 \* 27039 27138: gap of 100 bp  
 \* 27139 59886: contig of 31848 bp in length  
 \* 59887 59886: gap of 100 bp  
 \* 59887 164972: contig of 105886 bp in length.

## FEATURES

Location/Qualifiers

## source

1..164972

/organism="Mus musculus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10090"

/clone\_lib="RPCI-24 Male Mouse BAC"

1..50

/note="assembly\_fragment"

151..27038

/note="assembly\_fragment"

27139..59886

/note="assembly\_fragment"

59087..164972

/note="assembly\_fragment"

BASE COUNT 50238 a 31480 c 31671 g 51283 t 300 others

## ORIGIN

Query Match 10.8%; Score 38.2; DB 2; Length 164972;  
 Best Local Similarity 52.9%; Pred. No. 6.1;  
 Matches 82; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

Cy 2 GGAATTCGAGCAGATGAAGAAGCAGCGCTTGATTAATTCATAAGTTATTCGAGAGA 61  
 |||||  
 Db 43652 GAGAATTCAGGTTGATGAGAGAACTGATGTCATTATTTATCATCTACCTACCAAAA 43711  
 |||||

Cy 62 AAAAGGATTTACCGTCATTTCCGTTATTCACAAACGCGTGATTACGGAATTCCTCCAAAG 121  
 |||||  
 Db 43712 AATATGTTTCTCATTCATTCACCTTTCCATGAGATCTATATCATGATTCATCAGG 43771  
 |||||

Cy 122 CCGTAAAGATTTACGTTAATTCGAAACAGAAATAA 156  
 |||||  
 Db 43772 CCCCCAAGCATTCGCCACATCTAAACTAAAGAA 43806  
 |||||

## RESULT 15

AP003597 342850 bp DNA linear BCT 28-NOV-2001  
 LOCUS AP003597 Nostoc sp. PCC 7120 DNA, complete genome, section 17/19.  
 DEFINITION AP003597 BA000019  
 ACCESSION AP003597.1 GI:17133729  
 VERSION  
 KEYWORDS

SOURCE Nostoc sp. PCC 7120

ORGANISM Nostoc sp. PCC 7120

Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.

## REFERENCE

## AUTHORS

Kaneko, T., Nakamura, Y., Wolk, C.P., Kuritz, T., Sasamoto, S.,  
 Watanabe, A., Iriguchi, M., Ishikawa, A., Kawashima, K., Kimura, T.,  
 Kishida, Y., Kohara, M., Matsumoto, M., Matsuno, A., Muraki, A.,  
 Nakazaki, N., Shimpo, S., Sugimoto, M., Takazawa, M., Yamada, M.,  
 Yasuda, M. and Tabata, S.

Complete genomic sequence of the filamentous nitrogen-fixing

cyanobacterium Anabaena sp. strain PCC 7120

DNA Res. 8 (5), 205-213 (2001)

21595285

11759840

2 (bases 1 to 342850)

Direct Submission

Kaneko, T.

Submitted (02-MAY-2001) Takakazu Kaneko, Kazusa DNA Research

Institute, The First Laboratory for Plant Gene Research, Yana

1532-3, Kisarazu, Chiba 292-0812, Japan  
 (E-mail: kaneko@kazusa.or.jp,  
 URL: http://www.kazusa.or.jp/cyanobase/,  
 Tel: 81-438-52-3935 (ex.2338), Fax: 81-438-52-3934)

## FEATURES

source

1..342850

/organism="Nostoc sp. PCC 7120"

/mol\_type="genomic DNA"

/db\_xref="taxon:103690"

/note="synonym: Anabaena sp. PCC7120"

complement (174..1478)

/gene="all4593"

complement (174..1478)

/gene="all4593"

/note="ORF\_ID: all4593"

hypothetical protein"

/codon\_start=1

/transl\_table=11

/protein\_id="BAB76292.1"

/db\_xref="GI:17133730"

/translation="MDKSLIQLVDELPSDNI TVKVLNADLYVAPQWQNLVGFQDNTIR  
 AVTGETDARVQIKRDRATSLYNDPQQYQSAVKLYQITDKADTAMATAALANKYGEK  
 IGFLSFLNITPKADVTQSIDLVSKIAIEIIVFCKLNGIPQNPQPEFANSIANNYQDA  
 SLIRNVALVCLDGLPLGDFLSKIQISGAENSIITQNPVFLAVNSLPGSNPTDK  
 LCFINQGNFNAVQGMNLTGKITPQSISSHLNFIQIADNLDVFAAFIDQTNFYF  
 PHTGQSVARLILKAHTLVKEIKOETTOPIDQSSSAVKSDTSYAVSKIVEMDEE  
 FEDWYQATLEKWDOPFINVIGYSSNHWGKDDIETATYASSDANGYAVGKKVKC  
 WSDDEETWTATIEIKGNQFIHTGYDSSIDEVWNSDDIS"

1680..1880

/gene="asr4594"

1680..1880

/gene="asr4594"

/note="ORF\_ID: asr4594"

hypothetical protein"

/codon\_start=1

/transl\_table=11

/protein\_id="BAB76293.1"

/db\_xref="GI:17133731"

/translation="MTAIPQBSKPRQVLLYLGEDNYFVWEVPSLPGCCISQKTBREA  
 LTNIEVLQDKGFTLPEYQFINI"

1877..2770

/gene="alr4595"

1877..2770

/gene="alr4595"

/note="ORF\_ID: alr4595"

unknown protein"

/codon\_start=1

/transl\_table=11

/protein\_id="BAB76294.1"

/db\_xref="GI:17133732"

/translation="MSKLVDRLLLCFTFVLGFTTTHNVOAESAPFYWEFINVDIA  
 VOPNGDMLVTETQKYNFIRDVNHLSRVIPLDNVDRITEVSVENGQLLPSTISTG  
 NHQLLSNEHQLQSPESHTEVLKRVIGLVHNNYAOVYWEAIFSDRAKPAQOVR  
 VEPPELTNKKIIFOSFGVADIRVDTKTIEATVKQRIEPTGTLKIOVIFDRAQTKI  
 ETSQWQSTQFFYERLVWLLGLLVFGIFILLIIVSRNGTGTSSSIVGSSIDYVGGGG  
 DYGGGDYGGDYGGGGGGGGG"

3082..4473

/gene="alr4596"

3082..4473

/gene="alr4596"

/note="ORF\_ID: alr4596"

/codon\_start=1

/transl\_table=11

/product="malic enzyme"

/protein\_id="BAB76295.1"

/db\_xref="GI:17133733"

/translation="NADLTPTSFSVTLRLOIPNEVGLASITQAIASSGGNLGQIDL  
 IEQTRQVSLRDLTVDASTEHAETIVQAVKAPDKLYVVDRTENLRGGKISITSR  
 IFPKVSDDLMAATPGVGRVCATACQDSQVNLTKQNTVAIVDGGAVLGLNGLGP  
 HAALPYMECKAMLFKEFGLDAPFICLATQDTEETIIQAVNIDGVNLEDIAAPR  
 CEEI EKRLDELIDIPVHDDQGHATVTLAALFNKLVQKSLADIRIVINGAGAGV  
 ATARLLRKAGAELKWNCKSGIISTRTDLNEEKQEFVAKQATLAGAVQGVADFIGV  
 SAPGVLTPENWKSMAKDAIVFAMANPIPEIQELISKDVAVIATGCRSDYPNQINNVL

```

PPGVFRGALOCRAQTITTTMYLEAASAIASLVNPSDLSREHIIPSVFDERVVTAVAA
VQRAAREGARG"
5270. .6853
/gene="alr4597"
CDS
5270. .6853
/gene="alr4597"
/note="ORF_ID:alr4597"
/codon_start=1
/transl_table=11
/product="type I restriction-modification system DNA
methylase"
/protein_id="BAB76296.1"
/db_xref="GI:17133734"
/translation="MTBKAQKNSESFEOKLWKAADKLKRNIDAAYKHIVLGLIFL
KYSDAFEEHLQKLAGEGDYAGANPEDREYSAENVFFVPVEARWSYLQGAQKQPD
KTVLAMEAI EQNAKRLKGIIPKVYGOQKLDQKSLGGIDLGITTLGDAQAQD
VLGRVYVFLQOPALAEKGGQFYTPESIVKLLVEMLPYNGRVFDPCCGSGMFVQ
SEKRVKHQGLDDISYGOESNETTVYKLCRMNLAIRGIDGSIKNWPGSFLNDAHK
DLKADFYANPPNDSDMGELLNDRGMDKDLVPVGNANFAMWYSHETYLAPTGS
AGFVLSNGSSNNTSGGDIRKALVQKDLVDCIYMLPTQLPNTGIPACLNFLSRYKN
GNKRDHRGEVLPIDASELGYMNRSAFTEAIESKIADTYHEWKKSGSGYKDIKF
CKSAIAIEIKHNFVLPFGYVIGIPDEVEDGVSFEKMSLTWALGEQMRREGQLDEE
IKQLLVGFTVSDLVDF"
6858. .7121
/gene="asr4598"
6858. .7121
/gene="asr4598"
/note="ORF_ID:asr4598"
/unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAB76297.1"
/db_xref="GI:17133735"
/translation="MNTQVILKLNPIVIEKLUHQASHORTLEEEITSLIEDVTEN
TPITTSKRDWSGFFFEQTCAGWGELLVREPOFEAQEREPLL"
7226. .7804
/gene="alr4599"
7226. .7804
/gene="alr4599"
/note="ORF_ID:alr4599"
/hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAB76298.1"
/db_xref="GI:17133736"
/translation="MTAVKCKFKLPPEEYFIWEKKOOLPHEVINGEVYAMTGGTONH
GRASNIIFIVKHLRGCGQVGNDCRINIPETKYVPDVSTCDERORTAIQAIQ
YPLIVLVSFSTASYDRGKFRUYRNRNPSLDQDILLVDAEKIAIDLYKRNERNWEIF
NYQLGDNKIQSLDLSPIESYVEDIVFEESY"
7855. .8145
/gene="asr4600"
7855. .8145
/gene="asr4600"
/note="ORF_ID:asr4600"
/unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAB76299.1"
/db_xref="GI:17133737"
/translation="MIFKENMTQVILKQNLNPIVLEKLUHQASHORTLEEEITSLIE
DVAKEEPEKSSSGFMDNTIQFRMQQENITFDADFADIRDRSVGRDVEF"
8211. .8573
/gene="alr4601"
8211. .8573
/gene="alr4601"
/note="ORF_ID:alr4601"
/hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAB76300.1"
/db_xref="GI:17133738"
/translation="MVKKLTQHOLETATGVSVMHELLFGCLRLVESQKRRLLEYINQ
IPLKMTILNDLKAQWHAQCARLRSKMGKTPAFIDGQIASYSNNLILVTNNVSDF

```

```

ESFNDLAVENMFVNSGEG"
8577. .9749
/gene="alr4602"
8577. .9749
/gene="alr4602"
/note="ORF_ID:alr4602"
/codon_start=1
/transl_table=11
/product="type I site-specific deoxyribonuclease chain S"
/protein_id="BAB76301.1"
/db_xref="GI:17133739"
/translation="MSENKETTLCGLADIINGQSPGTGCTNNNGQGLPLNGPTFGD
RNPFTQITIDFKKIAAGDGLFCVRGTTGRMNWAKQYAGRGIASIRAKGILFCQ
PYIRAIIEKREKSLAVATGTFNFIKSHLNLIQVPSKNIKIYISNARIIDDKI
NLSRQNETLEAIQTLPFKWFIIDFEPNADGPKYSSGGAMVSALGYIPEAWSVGK
LQVYLNIRKHYAFKRGVEYITTEYTKILLTPVNFKIGGNDKSKYISADYSNEYVL
RRKDLAITMDLKEGDSLGYPAFIPDIKGVFLNORIGKVENNNIDKTFLYELLOR
REYSHILGTSSGSTRVHTSPSRICEYSFVIPDFELIDKESALATATIDKILLISTL"
9944. .10485
/gene="alr4603"
9944. .10485
/gene="alr4603"
/note="ORF_ID:alr4603"
/unknown protein"
/codon_start=1
/transl_table=11
/db_xref="GI:17133740"
/translation="MIGKRELNDERNFSFHIAALEESQABDKYNLLISKVIDLLFEV
KKLESISLIQYKNTLSNITDIRFLKFDGHSILKKNHDLIVDNNFIRTNIDNLNEYRLF"

Query Match      10.8%; Score 38.2; DB 1; Length 342850;
Best Local Similarity 55.7%; Pred. No. 6.1;
Matches 73; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY      34 TTGATAATTTTCATAAAGTTATTGGAAGAAAAAGAGTTTACCGTCCTTCGGTATTCACA 93
Db      243329 TAGATATTATTAAAGAAAAACATTGGAAGAACTTGGTTATTTTGTGGAAGTATGGCTTTAA 243388

QY      94 ATACGGCTGATTACGGAATTCGCCAAGCCGTAAGAGTTTACGTTAATTCGAACAGAA 153
Db      243389 ATGCTGCGGAATATGGCAATCCGCAAAATTAGAGAACGTAATTTATTGTTGGCAATAAA 243448

QY      154 TAACCAAGAA 164
Db      243449 AAGGTAAAGTA 243459

Search completed: November 15, 2003, 03:34:00
Job time : 1567.62 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 14, 2003, 23:55:19 ; Search time 227.827 Seconds  
(without alignments)  
4182.570 Million cell updates/sec

Title: US-09-928-457-83

Perfect score: 353

Sequence: 1 CGGAATCTGAGCAGATG.....ACGGAGGACCGTGGCTTT 353

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 510512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```

1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	353	100.0	353	19 AAV03591	Neisseria meningitidis
2	256.4	72.6	510	25 ABZ37757	N. gonorrhoeae nuc
3	256.4	72.6	510	25 ABZ41744	N. gonorrhoeae nuc
4	71	20.1	462	25 ABZ37756	N. gonorrhoeae nuc
5	71	20.1	462	25 ABZ41577	N. gonorrhoeae nuc
6	43.4	12.3	4519	19 AAV31192	E. coli J96 pathog
7	43.4	12.3	11165	21 ARA15186	DNA encoding Esche
8	35.6	10.1	13084	24 ABL34179	Human immune syste

C	9	35.4	10.0	935	21	AAZ80542	Human colon cancer
C	10	35.2	10.0	265118	22	AAH41227	Pyrococcus abyssi
C	11	34.2	9.7	1634	24	ABQ54282	Human ovarian anti
C	12	34	9.6	622	22	AAI17822	Human breast cancer
C	13	33.6	9.5	512	24	ABQ29518	Oligonucleotide fo
C	14	33.6	9.5	512	24	ABQ29519	Oligonucleotide fo
C	15	33.6	9.5	9742	24	ABL70479	Chemically treated
C	16	33	9.3	960	19	AAV53346	DNA encoding a Sta
C	17	33	9.3	2401	22	AAF56506	Arabidopsis CDC27A
C	18	33	9.3	2434	22	AAF56505	Arabidopsis CDC27A
C	19	32.6	9.2	14142	24	ABA01435	Streptococcus ther
C	20	32.4	9.2	421	22	AAK57045	Human immune/haema
C	21	32.4	9.2	828	21	AAH16500	Human colon cancer
C	22	32.2	9.1	2356	24	ABQ70511	Listeria monocytog
C	23	32.2	9.1	9155	24	ABL32463	Human immune syste
C	24	32	9.1	537	24	AET10194	Human breast cancer
C	25	32	9.1	580073	18	AAT58840	Mycoplasma genital
C	26	32	9.1	640681	24	ABA92787	Buchnera sp. genom
C	27	31.8	9.0	1564	21	AAAC98092	Human colon cancer
C	28	31.8	9.0	2195	22	AAH15162	Human cDNA sequenc
C	29	31.8	9.0	7434	20	AAK12971	Enterococcus faeca
C	30	31.8	9.0	7434	24	ABS98766	Enterococcus faeca
C	31	31.6	9.0	283	24	ABN75274	Human nuclease-lik
C	32	31.6	9.0	951	24	ABK78842	Bacillus clausii g
C	33	31.6	9.0	2854	24	ABL34346	Human immune syste
C	34	31.6	9.0	3312	25	ABZ58867	Human PPIP-180 po
C	35	31.6	9.0	4204	25	ABX70993	Novel human cDNA s
C	36	31.6	9.0	6064	25	ABX17516	Human cDNA encodin
C	37	31.6	9.0	14388	23	ABL05670	Drosophila melanog
C	38	31.6	9.0	82952	24	ABN85766	Arabidopsis yellow
C	39	31.6	9.0	1664976	19	AAV21209	Methanococcus jann
C	40	31.4	8.9	14160	22	AAK65583	Human immune/haema
C	41	31.4	8.9	42738	22	AAK68992	Human immune/haema
C	42	31.2	8.8	893	23	ABV30371	Human prostate exp
C	43	31.2	8.8	1504	22	AAH29913	C albicans apoptos
C	44	31.2	8.8	11394	24	ABK28222	DNA transcription
C	45	31.2	8.8	14924	24	ABL54322	Chemically treated

#### ALIGNMENTS

#### RESULT 1

AAV03591  
ID AAV03591 standard; DNA: 353 BP.

AC AAV03591;

XX 22-OCT-1998 (first entry)

DT Neisseria meningitidis DNA sequence C130.

DE N. gonorrhoeae; N. lactamica; chromosome 22491; region 1; region 2;

KW region 3; pathogenicity; blood-brain barrier; diagnosis; infection;

KW meningitis; ss.

XX Neisseria meningitidis.

OS Neisseria meningitidis.

PN WO9802547-A2.

XX 22-JAN-1998.

FF 11-JUL-1997; 97WO-FR01295.

XX 12-JUL-1996; 96PR-0008768.

PR (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.

PA (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.

PA (SMIK ) SMITHKLINE BEECHAW.

PI Nassif X, Tinsley C, Achtman M, Merker P, Ruelle J;

PI Vinals C;

DR WPI; 1998-110594/10.  
 XX Genes present in *Neisseria meningitidis* but not other *Neisseria*  
 PT species - and related host cells, RNA, anti-sense sequences,  
 PT polypeptide(s) and antibodies, useful for diagnosing *Neisseria*  
 PT meningitidis infection and in protective vaccines  
 XX  
 PS Example 4; Pages 121-122; 150pp; French.  
 XX  
 CC AAV03575-606 represent sequences that are present in *Neisseria*  
 CC meningitidis and *N. gonorrhoeae* but not in *N. lactamica*, except for the  
 CC genes involved in biosynthesis of the capsule polysaccharide, *frpA* or *C*,  
 CC *opc*, *porA*, *rotamase*, sequence IC1106, *IgA* protease, *pillin*, *pilC*,  
 CC proteins which bind transferrin and opacity proteins. The DNA sequences  
 CC are responsible for the differences in pathogenicity between *N.*  
 CC meningitidis and *N. gonorrhoeae*, specifically they include the genes that  
 CC allow *N. meningitidis* to cross the blood-brain barrier. DNA sequences  
 CC *lactamica*, are responsible for colonisation and penetration of the  
 CC mucosa. The DNA sequences can be used to produce probes and primers, and  
 CC antibodies produced against the encoded proteins are used in standard  
 CC hybridisation/immunoassay processes for diagnosis of *N. meningitidis*  
 CC infection, particularly meningitis.  
 XX  
 SQ Sequence 353 BP; 119 A; 63 C; 82 G; 89 T; 0 other;  
 Query Match 100.0%; Score 353; DB 19; Length 353;  
 Best Local Similarity 100.0%; Pred. No. 6.7e-98;  
 Matches 353; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CGGGAATTCGAGCAGATGAAGAAAGCAGGCTTGATAATTCATAAAGTTATGGAAG 60  
 DB 1 CGGGAATTCGAGCAGATGAAGAAAGCAGGCTTGATAATTCATAAAGTTATGGAAG 60  
 QY 61 AAAAAGGATTTACGTCATTCGGTATTCACATACGGCTGATTACGGAATCCCCAAA 120  
 DB 61 AAAAAGGATTTACGTCATTCGGTATTCACATACGGCTGATTACGGAATCCCCAAA 120  
 QY 121 GCCGTAAAGATTTACGTTAATTCGAAACAGAAATTAACAAAGAAAGCTGGAACCACTCA 180  
 DB 121 GCCGTAAAGATTTACGTTAATTCGAAACAGAAATTAACAAAGAAAGCTGGAACCACTCA 180  
 QY 181 AGTATTCGGGCAACCGCTTACGGTAGCCGATGTTTGGGAATGGAATGGCTTCCCAA 240  
 DB 181 AGTATTCGGGCAACCGCTTACGGTAGCCGATGTTTGGGAATGGAATGGCTTCCCAA 240  
 QY 241 CATTATTCGAGACACCAAGACGAAACGGATTTTATGCATAGCTGTCGGGAATTATCTG 300  
 DB 241 CATTATTCGAGACACCAAGACGAAACGGATTTTATGCATAGCTGTCGGGAATTATCTG 300  
 QY 301 ATATCACTTGAACGATTGGCTTGATCACTTAAACCGGAGAACCGTTGGGCTT 353  
 DB 301 ATATCACTTGAACGATTGGCTTGATCACTTAAACCGGAGAACCGTTGGGCTT 353  
 RESULT 2  
 ABZ37757  
 ID ABZ37757 standard; DNA; 510 BP.  
 XX  
 AC ABZ37757;  
 XX  
 DT 07-MAR-2003 (first entry)  
 XX  
 DE *N. gonorrhoeae* nucleotide sequence SEQ ID 103.  
 XX  
 KW Antibacterial; infection; vaccine; gene therapy; gene; ds.  
 XX  
 OS *Neisseria gonorrhoeae*.  
 XX  
 PN WO200279243-A2.  
 XX  
 XX 10-OCT-2002.  
 XX

PF 12-FEB-2002; 2002WO-IB02069.  
 XX  
 PR 12-FEB-2001; 2001GB-0003424.  
 XX  
 PA (CHIR-) CHIRON SPA.  
 XX  
 PI Fontana MR, Pizza M, Massignani V, Monaci E;  
 XX  
 DR WPI; 2003-058415/05.  
 DR P-PSDB; ABP76787.  
 XX  
 PT New protein from *Neisseria gonorrhoeae*, useful for the manufacture of a  
 PT medicament for treating or preventing *N. gonorrhoeae* infection -  
 XX  
 PS Claim 6; Page 189; 815pp; English.  
 XX  
 CC The present invention relates to proteins from *Neisseria gonorrhoeae*.  
 CC Also disclosed are the nucleic acid molecules encoding the proteins and  
 CC antibodies that specifically bind to the proteins. The composition  
 CC comprising the protein, nucleic acid or antibody is useful for the  
 CC manufacture of a medicament for treating or preventing *N. gonorrhoeae*  
 CC infection, this may be in the form of a vaccine or gene therapy.  
 CC Sequences given in records ABZ37706-ABZ42016 represent nucleic acid  
 CC molecules of the invention.  
 XX  
 SQ Sequence 510 BP; 173 A; 80 C; 117 G; 140 T; 0 other;  
 Query Match 72.6%; Score 256.4; DB 25; Length 510;  
 Best Local Similarity 95.5%; Pred. No. 2.5e-68;  
 Matches 236; Conservative 0; Mismatches 11; Indels 3; Gaps 3;  
 QY 1 CGGGAATTCGAGCAGATGAAGAAAGCAGGCTTGATAATTCATAAAGTTATGGAAG 60  
 DB 203 CGGGAATTCGAGCAGATGAAGAAAGCAGGCTTGATAATTCATAAAGTTATGGAAG 262  
 QY 61 AAAAAGGATTTACGTCATTCGGTATTCACATACGGCTGATTACGGAATCCCCAAA 120  
 DB 263 AAAAAGGATTTACGTCATTCGGTATTCACATACGGCTGATTACGGAATCCCCAAA 322  
 QY 121 GCCGTAAAGATTTACGTTAATTCGAAACAGAAATTAACAAAGAAAGCTGGAACCACTCA 180  
 DB 323 GCCGTAAAGATTTACGTTAATTCGAAACAGAAATTAACAAAGCTGGAACCACTCA 381  
 QY 181 AGTATTCGGGCAACCGCTTACGGTAGCCGATGTTTGGGAATGGAATGGCTTCCCAA 239  
 DB 382 AGTATTCGGGCAACCGCTTACGGTAGCCGATGTTTGGGAATGGAATGGCTTCCCAA 441  
 QY 240 ACATTTATTCGAGACACCAAGACGAAACGGATTTTATGCATAGCTGTCGGGAATTATCT 299  
 DB 442 AAATTTATTCGAGACACCAAGACGAAACGGATTTTATGCATAGCTGTCGGGAATTATCT 500  
 QY 300 GATATCACTT 309  
 DB 501 GATATCACTT 510  
 RESULT 3  
 ABZ41744  
 ID ABZ41744 standard; DNA; 510 BP.  
 XX  
 AC ABZ41744;  
 XX  
 DT 07-MAR-2003 (first entry)  
 XX  
 DE *N. gonorrhoeae* nucleotide sequence SEQ ID 8077.  
 XX  
 KW Antibacterial; infection; vaccine; gene therapy; gene; ds.  
 XX  
 OS *Neisseria gonorrhoeae*.  
 XX  
 PN WO200279243-A2.  
 XX  
 XX 10-OCT-2002.  
 XX

XX 12-FEB-2002; 2002WO-IB02069.  
PF XX  
XX 12-FEB-2001; 2001GB-0003424.  
PR XX  
XX (CHIR-) CHIRON SPA.  
PA XX  
XX Fontana MR, Pizza M, Masignani V, Monaci E;  
PI XX  
XX WPI; 2003-058415/05.  
DR XX  
DR P-PSDB; ABP80774.  
XX  
XX New protein from Neisseria gonorrhoeae, useful for the manufacture of a  
PT medicament for treating or preventing N. gonorrhoeae infection -  
PT medicament for treating or preventing N. gonorrhoeae infection -  
XX  
XX Disclosure; Page 778; 815pp; English.  
PS  
XX The present invention relates to proteins from Neisseria gonorrhoeae.  
CC Also disclosed are the nucleic acid molecules encoding the proteins and  
CC antibodies that specifically bind to the proteins. The composition  
CC comprising the protein, nucleic acid or antibody is useful for the  
CC manufacture of a medicament for treating or preventing N. gonorrhoeae  
CC infection, this may be in the form of a vaccine or gene therapy.  
CC Sequences given in records AB237706-AB242016 represent nucleic acid  
CC molecules of the invention.  
XX  
SQ Sequence 510 BP; 173 A; 80 C; 117 G; 140 T; 0 other;  
Query Match 72.6%; Score 256.4; DB 25; Length 510;  
Best Local Similarity 95.5%; Pred. No. 2.5e-68;  
Matches 296; Conservative 0; Mismatches 11; Indels 3; Gaps 3;  
QY 1 CGGGAATCTGAGCAGATGAAGAAGCAGCGCTTGATTAATTCATAAAGTTATTGGAAG 60  
Db 203 CGGGAATCTGAGCAGATGAAGAAGCAGCGCTTGATTAATTCATAAAGTTATTGGAAG 262  
QY 61 AAAAGAGATTACCGTCCATTCGGTATTCACATACGGCTGATACGGATTCCCAAA 120  
Db 263 AAAAGAGATTACCGTCCATTCGGTATTCACATACGGCTGATACGGATTCCCAAA 322  
QY 121 GCCGTAAAGATTACGTTAATTCGAAACAGAAATACCAAGAAAGCTGGAACCACTCA 180  
Db 323 GCCGTAAAGATTACGTTAATTCGAAACAGAAATACCAAGAAAGCTGGAACCACTCA 381  
QY 191 AGTATTCGGCAACCGCTTACGGTAGCCGATGTTTGGGAATGG-AAATGGCTTTCCCA 239  
Db 382 AGTATTCGGCAACCGCTTACGGTAGCCGATGTTTGGGAATGGAAACCGCTTTCCCA 441  
QY 240 ACATTATTGCAGGACACCAACAGCAAGAAACGGATTTTATGCATAGCTGTGCGGATTTACT 299  
Db 442 AAATTATGCGAGGACATCAAGACGAGACGGATTTTATGCATAGTTGTGCGGG-ATTATCC 500  
QY 300 GATATCACTT 309  
Db 501 GATATCAATT 510  
RESULT 4  
AB237756  
ID AB237756 standard; DNA; 462 BP.  
XX  
AC  
XX AB237756;  
XX  
DT 07-MAR-2003 (first entry)  
XX  
DE N. gonorrhoeae nucleotide sequence SEQ ID 101.  
XX  
XX Antibacterial; infection; vaccine; gene therapy; gene; ds.  
XX  
XX Neisseria gonorrhoeae.  
XX  
EN WO200279243-A2.

PD 10-OCT-2002.  
XX  
PF 12-FEB-2002; 2002WO-IB02069.  
XX  
PR 12-FEB-2001; 2001GB-0003424.  
XX  
XX (CHIR-) CHIRON SPA.  
PA  
XX Fontana MR, Pizza M, Masignani V, Monaci E;  
PI  
XX WPI; 2003-058415/05.  
DR  
DR P-PSDB; ABP76786.  
XX  
XX New protein from Neisseria gonorrhoeae, useful for the manufacture of a  
PT medicament for treating or preventing N. gonorrhoeae infection -  
PT medicament for treating or preventing N. gonorrhoeae infection -  
XX  
XX Claim 6; Page 188; 815pp; English.  
PS  
XX The present invention relates to proteins from Neisseria gonorrhoeae.  
CC Also disclosed are the nucleic acid molecules encoding the proteins and  
CC antibodies that specifically bind to the proteins. The composition  
CC comprising the protein, nucleic acid or antibody is useful for the  
CC manufacture of a medicament for treating or preventing N. gonorrhoeae  
CC infection, this may be in the form of a vaccine or gene therapy.  
CC Sequences given in records AB237706-AB242016 represent nucleic acid  
CC molecules of the invention.  
XX  
SQ Sequence 462 BP; 135 A; 89 C; 114 G; 124 T; 0 other;  
Query Match 20.1%; Score 71; DB 25; Length 462;  
Best Local Similarity 85.7%; Pred. No. 1.1e-11;  
Matches 114; Conservative 0; Mismatches 15; Indels 4; Gaps 3;  
QY 215 TTGGGAATGGAATGCGCTTCCACATATTGCGAGACACCAAGACGAGCGGATTT 274  
Db 1 TTGGGAATGGAATGCGCTTCCACATATTGCGAGACACCAAGACGAGCGGATTT 60  
QY 275 ATCCATAGCTGTGCGGAATTATCTGATATCACTT--GAACGATTGGC-TTGATACCTAA 331  
Db 61 ATCCATAGCTGTGCGGG-ATTATCCGATATCACTTGAACGATTGGCTTGATACCTAA 119  
QY 332 AAACGGAGGAAC 344  
Db 120 AAACGGAGGAAC 132  
RESULT 5  
AB241577  
ID AB241577 standard; DNA; 462 BP.  
XX  
AC AB241577;  
XX  
DT 07-MAR-2003 (first entry)  
XX  
DE N. gonorrhoeae nucleotide sequence SEQ ID 7743.  
XX  
XX Antibacterial; infection; vaccine; gene therapy; gene; ds.  
XX  
XX Neisseria gonorrhoeae.  
XX  
EN WO200279243-A2.  
XX  
DT 10-OCT-2002.  
XX  
PF 12-FEB-2002; 2002WO-IB02069.  
XX  
PR 12-FEB-2001; 2001GB-0003424.  
XX  
XX (CHIR-) CHIRON SPA.  
PA  
XX Fontana MR, Pizza M, Masignani V, Monaci E;  
PI  
XX WPI; 2003-058415/05.

DR P-PSDB; ABP80607.

XX New protein from *Neisseria gonorrhoeae*, useful for the manufacture of a

PT medicament for treating or preventing *N. gonorrhoeae* infection -

XX

XX Disclosure; Page 754; 815pp; English.

XX The present invention relates to proteins from *Neisseria gonorrhoeae*.

CC Also disclosed are the nucleic acid molecules encoding the proteins and

CC antibodies that specifically bind to the proteins. The composition

CC comprising the protein, nucleic acid or antibody is useful for the

CC manufacture of a medicament for treating or preventing *N. gonorrhoeae*

CC infection, this may be in the form of a vaccine or gene therapy.

CC Sequences given in records AB237706-AB242016 represent nucleic acid

CC molecules of the invention.

XX

SQ Sequence 462 BP; 135 A; 89 C; 114 G; 124 T; 0 other;

Query Match 20.1%; Score 71; DB 25; Length 462;

Best Local Similarity 85.7%; Pred. No. 1.1e-11;

Matches 114; Conservative 0; Mismatches 15; Indels 4; Gaps 3;

QY 215 TTGGGAATGGAATGCGTTTCCCAACATTTTCGCGACACCAAGACGAAACGGATTTT 274

DB 1 TTGGGAATGGAATGCGTTTCCCAACATTTTCGCGACACCAAGACGAAACGGATTTT 60

QY .275 ATGCATAGCTGCGGGGAATTTCTGATATCACTT--GAACGATTGGC-TTGATACCTAA 331

DB 61 ATGCATAGTTGCGGG-ATTATCGATATCAATTGAAACGATTGCTTTGATACCTAA 119

QY 332 AAACGGAGCAACC 344

DB 120 AAACGGAGCAAAC 132

RESULT 6

AAV31192/c

ID AAV31192 standard; DNA; 4519 BP.

XX

AC AAV31192;

DT 01-OCT-1998 (first entry)

XX

DE *E. coli* J96 pathogenicity island contig #6.

XX

XX PAI; pathogenicity island; uropathogenic *E. coli* detection; PAI IV; pheR;

KW PAI V; pheV; vaccine; protective immune response; ds.

XX

OS *Escherichia coli*.

XX

XX WO9822575-A2.

PN

PD 28-MAY-1998.

XX

XX 21-NOV-1997; 97WO-US21347.

PF

XX 14-OCT-1997; 97US-0061953.

PR

XX 22-NOV-1996; 96US-0031626.

PR

XX (HUMA-) HUMAN GENOME SCI INC.

PA

XX (UYWI-) UNIV WISCONSIN.

PA

XX Choi GH, Dillon PJ, Welch RA;

PI

XX WPI; 1998-312461/27.

DR

XX New isolated uropathogenic *E. coli* nucleotide sequences - used to

PT develop products for the detection of pathogenic *E. coli* and to

PT elicit an immune response to pathogenic *E. coli*

XX

PS Claim 21; Page 94-96; 250pp; English.

XX

CC This sequence represents a *E. coli* strain J96 contig containing

CC pathogenicity island (PAI) sequences, and represents a nucleic acid

CC molecule of the invention. PAIs are large fragments of DNA which comprise

CC pathogenicity determinants. The sequences of the invention are taken from

CC PAI IV and PAI V. PAI IV is located at approximately 64 min (near pheV)

CC on the *E. coli* chromosome and is greater than 170 kb. PAI V is located at

CC approximately 94 min (at pheR) on the *E. coli* chromosome and is

CC approximately 160 kb in size. Antibodies specific to the proteins encoded

CC by the PAI open reading frames of the invention can be used in kits to

CC detect uropathogenic *E. coli*. The proteins are used in vaccines to elicit

CC a protective immune response in an animal to the uropathogenic *E. coli*

CC strain J96.

XX

SQ Sequence 4519 BP; 1141 A; 1101 C; 918 G; 1352 T; 7 other;

Query Match 12.3%; Score 43.4; DB 19; Length 4519;

Best Local Similarity 55.7%; Pred. No. 0.0064;

Matches 83; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 21 AAAGAAGCAGCGTTTCATAATTTCATAAAGTTATTGGAGAAAGGATTTACCGTCCAT 80

DB 979 AAAGAAGCAGCGTTTCAGAGCTTTATTAGCTACTTAAAGAGTTAGAGTAACTATA 920

QY 81 TTCCGTATTACATACGCTGATTACGGAATTTCCCAAGCCGTAAAGATTTTACGTTA 140

DB 919 TCTTTTAGCCAAATGCTGAGAACTATGGGATTTCCCAAGAAAGAAAGACTCGTGCTC 860

QY 141 ATTGCAACAGAAATACCAAGAAAGAAAGCT 169

DB 859 TTAGTAGTCGAGTAGTAGTAAAGTTACCT 831

RESULT 7

AAAI5186

ID AAI5186 standard; DNA; 11165 BP.

XX

AC AAI5186;

XX

DT 04-SEP-2000 (first entry)

XX

DE DNA encoding *Escherichia coli* virulence proteins.

XX

XX Virulence protein; tatA; tatB; tatC; tatE; mdoG; creC; recG; yggN;

KW eck1; iroD; iroC; iroE; mtd2; msl; vaccine; infection;

KW Gram negative bacterium; ss.

XX

OS *Escherichia coli*.

XX

XX Key Location/Qualifiers

FF CDS 2..1099

FT /tag= a

FT /product= "virulence protein"

FT /note= "encodes AAY93237; no termination codon given"

FT 1102..1488

FT /tag= b

FT /product= "virulence protein"

FT /note= "encodes AAY93238"

FT 1573..1896

FT /tag= c

FT /product= "virulence protein"

FT /note= "encodes AAY93239"

FT 1939..2196

FT /tag= d

FT /product= "virulence protein"

FT /note= "encodes AAY93240; no termination codon given"

FT 2198..2533

FT /tag= e

FT /product= "virulence protein"

FT /note= "encodes AAY93241"

FT 2613..3041

FT /tag= f

FT /product= "virulence protein"

FT /note= "encodes AAY93242; no termination codon given"

FT 3054..3410

FT CDS





Best Local Similarity 54.6%; Pred. No. 2.3;  
Matches 71; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 37 ATAAATTTTCAAAAGTTATTGCAAGCAAAAGGATTACCGTCCATTTCGGTATTACAAATA 96  
DB 9929 AAATATTCAATAAATATACAAATAAATAAATAAATAAATAAATAAATATCTATCATACAA 9870

QY 97 CGGCTGATTACGGAATTCCTCCAAAGCCGTAAAGATTACGTTAAATTCGCAACAGATAA 156  
DB 9869 TCAATATACAAACACCCCTTAATAACCAAAACATAAACTTTATCTCTCTAAATAA 9810

QY 157 CCAAGAAAA 166  
DB 9809 CCAAAAAAAA 9800

RESULT 9  
AAZ80542/c  
ID AAZ80542 standard; cDNA; 935 BP.  
XX  
AC AAZ80542;  
XX  
DT 07-APR-2000 (first entry)  
XX  
DE Human colon cancer cell line SW480 cDNA clone SEQ ID NO:626.  
XX  
KW Human; gene expression product; diagnosis; tumour; colon cancer;  
KW colorectal adenocarcinoma; cell line SW480; cell proliferation;  
KW cytostatic; sarcoma; breast cancer; neoplasia; dysplasia;  
KW hyperplasia; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO9964576-A2.  
XX  
PD 16-DEC-1999.  
XX  
PF 09-JUN-1999; 99WO-IB01062.  
XX  
PR 10-JUN-1999; 98US-0089801.  
XX  
PA (FARB ) BAYER CORP.  
XX  
PI Endege WO, Steinmann KE, Astle JH, Burgess CC, Bushnell SE;  
PI Carroll E, Catino TJ, Derti A, Ford DM, Lewis ME, Monahan JE;  
PI Schlegel R;  
XX  
DR WPI; 2000-087220/07.  
XX  
CC Novel nucleic acids, used to develop products for the diagnosis and  
PT treatment of disorders involving unwanted cell proliferation,  
PT particularly cancers, especially colon cancer.  
XX  
PS Claim 15; Page 379; 469pp; English.  
XX  
CC AAZ79917 to AAZ80766 represent double stranded cDNA clones isolated from  
CC the human colorectal adenocarcinoma (colon cancer) cell line SW480. The  
CC cDNA clones can be used to generate antisense oligonucleotides which  
CC can be used for antisense therapy. Methods and products from the present  
CC invention can be used for identifying and/or classifying cancerous cells  
CC present in a human tumour, particularly in solid tumours, e.g.  
CC carcinomas and sarcomas, e.g. breast or colon cancers. The cDNA clones  
CC can be used for developing agents for the diagnosis and treatment of  
CC disorders involving unwanted cell proliferation, such as neoplasia,  
CC dysplasia or hyperplasia.  
XX  
SQ Sequence 935 BP; 216 A; 216 C; 236 G; 211 T; 56 other;

Query Match 10.0%; Score 35.4; DB 21; Length 935;  
Best Local Similarity 50.0%; Pred. No. 1;  
Matches 57; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 129 AGATTACCTTAATTCGCAACAGATAACCAAGAAAGCTGCAACCACTCAAGTATTCG 188

DB 342 AAATTTTNNCCACCCCAAGGAAAMCCAAAGGCTTAATGGACCAATGATTTCAG 283  
QY 189 GCACAACGGCTTACGTTAGCCGATGTTTGGGAAT 223  
DB 282 GCCCAAGGCTTGGGTACCCGAGCTTCGGGATT 248

RESULT 10  
AAH41227/c  
ID AAH41227 standard; DNA; 265118 BP.  
XX  
AC AAH41227;  
XX  
DT 29-OCT-2001 (first entry)  
XX  
DE Pyrococcus abyssi genomic fragment #6.  
XX  
KW Hyperthermophilic archaeon; hyperthermophilic protein; ds.  
XX  
OS Pyrococcus abyssi.  
XX  
FH Key Location/Qualifiers  
FT misc\_feature 1..49980  
FT /tag= a  
FT /note= "This sequence overlaps with the 3' end of  
FT AAH41226"  
XX  
PN FR2792651-A1.  
XX  
PD 27-OCT-2000.  
XX  
PF 21-APR-1999; 99FR-0005034.  
XX  
PR 21-APR-1999; 99FR-0005034.  
XX  
PA (CNRS ) CNRS CENT NAT RECH SCI.  
PA (IFRE-) IFREMER INST FR RECH EXPL MER.  
XX  
PI Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;  
PI Querellou J, Weissenbach J, Saurin W, Heilig R;  
XX  
DR WPI; 2001-126236/14.  
XX  
CC New nucleotide sequences isolated from Pyrococcus abyssi encode  
PT proteins useful in industry.  
XX  
PS Claim 1; Page 593-665; 1657pp; French.  
XX  
CC The present invention relates to the genomic sequence of Pyrococcus  
CC abyssi and P. abyssi proteins (see AAB96053-AAB96842). P. abyssi is a  
CC hyperthermophilic archaeon, which is isolated from deep-sea hydrothermal  
CC vents. The present sequence is a fragment of the genomic sequence of P.  
CC abyssi. The 5' end of this sequence overlaps with the 3' end of AAH41226.  
CC The proteins of the present invention have various potential industrial  
CC uses, since the proteins are stable at very high temperatures, some up to  
CC 110 degrees centigrade.  
CC Note: This patent is in the same patent family as WO2000085062, which  
CC contains additional sequences as shown in AAB99132-AAB99143,  
CC AAB75903-AAH75920 and AAG66436.  
XX  
SQ Sequence 265118 BP; 75226 A; 61302 C; 54005 G; 74585 T; 0 other;

Query Match 10.0%; Score 35.2; DB 22; Length 265118;  
Best Local Similarity 52.8%; Pred. No. 8.5;  
Matches 76; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 112 TTCCCCAAAGCCGTAAGATTTACGTTAATTCGCAACAGATAACCAAGAAAGCTCG 171  
DB 91600 TACCCCAAGGATTGCAAGGGCCATGGTAAATCTAACGAGGCAACTAGGAAATACTAG 91541  
QY 172 AACCACTCAAGTATTCGGGCAACGGCTTACCGTACCCGATGTTTGGAAATGGAATCG 231

Db 91540 ATCTTTCATGGGTACAGGGGAATGCTTATGAGGCTGGTTTAAATGGGCTGAAAGTTT 91491  
 QY 232 CTTTCCCAACATTATGCGAGGACA 255  
 Db 91480 ATGGAATAGATATAAGGGGAGGACA 91457

RESULT 11  
 ABQ54282/c  
 ID ABQ54282 standard; cDNA; 1634 BP.

XX ABQ54282;

XX 22-AUG-2002 (first entry)

XX DE Human ovarian antigen HNEUA49 cDNA, SEQ ID NO:162.

XX KW Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;  
 KW ovarian cancer; breast cancer; tumour; reproductive system disorder;  
 KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;  
 KW PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; infection;  
 KW inflammatory condition; immune disorder; blood disorder;  
 KW cardiovascular disorder; respiratory disorder; neurological disorder;  
 KW gastrointestinal disorder; urinary system disorder; drug screening;  
 KW gene therapy; chromosome mapping; forensic analysis;  
 KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;  
 KW antiinflammatory; gynaecological; reproductive; gene; ss.

XX OS Homo sapiens.

XX PN WO200200677-A1.

XX PD 03-JAN-2002.

XX PF 07-JUN-2001; 2001WO-US18569.

XX PR 07-JUN-2000; 2000US-209467P.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Birse CE, Rosen CA;

XX WPI; 2002-147878/19.

XX DR P-PSDB; ABP41205.

XX Isolated nucleic acid molecules encoding novel ovarian polypeptides,  
 PT useful in the prevention, treatment and diagnosis of cancer (e.g.  
 PT ovarian cancer), immune disorders, cardiovascular disorders and  
 PT neurological diseases -

XX PS Claim 1; SEQ ID No 162; 2922pp; English.

XX The invention relates to 2175 novel human ovarian antigens (ABP41054-  
 CC ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also  
 CC encompasses polypeptides 90% identical and polynucleotides 95% identical  
 CC to the sequences of the invention. The invention additionally relates to  
 CC recombinant vectors and host cells comprising human ovarian antigen  
 CC polynucleotides, antibodies against human ovarian antigens and the use  
 CC of ovarian antigen polynucleotides and polypeptides in diagnosing,  
 CC treating, prognosing or preventing various ovary and/or breast-related  
 CC disorders. Such conditions include ovarian cancer and breast cancer, and  
 CC metastatic tumours of ovarian or breast origin, reproductive system  
 CC disorders (e.g., infertility, disorders of pregnancy, anovulation,  
 CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine  
 CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic  
 CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and  
 CC vaginitis), immune disorders (e.g., congenital and acquired  
 CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),  
 CC blood-related disorders (e.g., anaemia), cardiovascular disorders,  
 CC respiratory disorders, neurological disorders, gastrointestinal disorders  
 CC and urinary system disorders. Ovarian antigen polypeptides and  
 CC polynucleotides may also be used in screening for compounds which  
 CC modulate ovarian antigen expression or activity. The polynucleotides may

CC further be used for gene therapy, chromosome mapping, in the  
 CC identification of individuals and in forensic analysis, and the  
 CC polypeptides may be used as food additives or to prepare antibodies  
 CC useful in disease diagnosis, drug targeting and phenotyping. The present  
 CC sequence represents cDNA encoding a human ovarian antigen of the  
 CC invention.

CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 1634 BP; 393 A; 403 C; 351 G; 473 T; 14 other;

Query Match 9.7%; Score 34.2; DB 24; Length 1634;  
 Best Local Similarity 50.0%; Pred. No. 2.9;  
 Matches 81; Conservative 1; Mismatches 80; Indels 0; Gaps 0;

QY 24 GAAAGCAGGCTTGATAATTTCAATAAGTTATTGGAGAAAAAGGATTTACCGTCCATTTC 83  
 Db 1631 GNACGGTGGGTTTWTITTTTTTTTTTTTTTTTTTTTAAAAAGGTGGACTTGAATGTT 1572  
 QY 84 GGTATTACACAATACGGGTGATTACGGAATTCCTCCAAAGCCGTAAAGATTTACGTTAATT 143  
 Db 1571 TATTACAGGATGCTGCAAGATAGGAATTCACATAGAAATTAGAAACCTAGTCAGAGG 1512  
 QY 144 GCAACAGAAATAACCAAGAAAAAGCTGGAAACCAAGTCAAGTAT 185  
 Db 1511 ACAAGCTTCATACAGTATGTACAGTTGGAACTGTTCAAGTAT 1470

RESULT 12

AAAL17822/c

ID AAL17822 standard; cDNA; 622 BP.

XX AC AAL17822;

XX DT 07-DEC-2001 (first entry)

XX DE Human breast cancer expressed polynucleotide 10279.

XX KW Human; breast cancer; cell marker; cytostatic; ss.

XX OS Homo sapiens.

XX PN WO200151628-A2.

XX PD 19-JUL-2001.

XX PF 10-JAN-2001; 2001WO-US00798.

XX PR 14-JAN-2000; 2000US-0176077.

XX PR 14-MAR-2000; 2000US-0189167.

XX PR 24-MAR-2000; 2000US-0192099.

XX PR 29-MAR-2000; 2000US-0193480.

XX PR 15-MAY-2000; 2000US-0205230.

XX PR 09-JUN-2000; 2000US-0211315.

XX PR 25-JUL-2000; 2000US-0220534.

XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX PI Lillie J, Xu Y, Wang Y, Steinmann K;

XX WPI; 2001-451856/48.

XX New peptide useful as a marker for the diagnosis of breast cancer -

XX PS Claim 1; Page 1832-1833; 3695pp; English.

XX The invention relates to human breast cancer expressed polynucleotides  
 CC (AAL07544-AAL26789) and methods of assessing whether a patient is  
 CC afflicted with breast cancer by examining the correlation between the  
 CC expression of certain markers and the cancerous state of breast cells.  
 CC The polynucleotides and encoded polypeptides are potential markers for  
 CC detecting, diagnosing, monitoring, characterising treating and

CC potentially preventing breast cancer. The polynucleotides and encoded  
 CC polypeptides are also useful for isolating compounds with cytostatic  
 CC activity.  
 XX  
 SQ Sequence 622 BP; 177 A; 91 C; 77 G; 277 T; 0 other;  
 Query Match 9.6%; Score 34; DB 22; Length 622;  
 Best Local Similarity 50.0%; Pred. No. 2.4; Indels 0; Gaps 0;  
 Matches 85; Conservative 0; Mismatches 85;  
 QY 80 TTTCGGTATTACCAATACGGCTGATTACGGAAATCCCAAGCGGTAAAGATTACGTT 139  
 DB 493 TTTCGGTATTAAAGGCAATACGGAAGGATTAACCAACAAATGTTTAAATC 424  
 QY 140 AATTGCAACAGATTAACCAAGAAAGCTGGACCAAGTCAAGTATTCGGGCAACGGCT 199  
 DB 423 ATTTTTATGAAGAGAAACAGGTAAAGTTCAATTCAATCAAGATGTCAGTTAGCCATA 364  
 QY 200 TACGTACCGCATGTTTTCGGAAATGGAATGGCTTCCCAACATTATGC 249  
 DB 363 TTGACTAACCAACTCTGTGACAGCTCTTCTACTGCCAAATTAATGC 314

RESULT 13  
 ABO29518/c  
 ID ABO29518 standard; DNA; 512 BP.  
 XX  
 AC ABO29518;  
 XX  
 DT 12-JUL-2002 (first entry)  
 XX  
 DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 16109.  
 KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
 KW drug; side effect; cancer; central nervous system; cardiovascular;  
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
 KW SNP; cell differentiation; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200218632-A2.  
 XX  
 PD 07-MAR-2002.  
 XX  
 PF 01-SEP-2001; 2001WO-EP10074.  
 XX  
 PR 01-SEP-2000; 2000DE-1043826.  
 PR 05-SEP-2000; 2000DE-1044543.  
 XX  
 PA (EPIG-) EPIGENOMICS AG.  
 XX  
 PI Olek A, Piepenbrock C, Berlin K, Guetig D;  
 XX  
 DR WPI; 2002-371829/40.  
 XX  
 PT Determining the degree of cytosine methylation in genomic DNA, useful  
 PT for diagnosis and prognosis, comprises selective hybridization of  
 PT amplicons from chemically treated DNA  
 XX  
 PS Claim 12; 56pp + Sequence Listing; 56pp; German.

XX This invention describes a novel method for determining the degree of  
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
 CC genomic sample of DNA. The sample is treated chemically to convert  
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
 CC DNA that contains the target C is amplified to form a labeled amplicon.  
 CC The amplicon is hybridised to two classes, each with at least one  
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
 CC and the degree of hybridisation to both classes is determined from the  
 CC label on the amplicon. From the ratio of labels hybridised to the two  
 CC classes of oligomers, the degree of methylation is calculated. The method  
 CC is used: (i) for diagnosis and/or prognosis of side effects of  
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders

CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
 CC systems etc., particularly by detecting mutations or single nucleotide  
 CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue  
 CC types and for investigating cell differentiation. The method allows the  
 CC methylation status of many C residues to be determined simultaneously.  
 CC ABO29518-ABQ54121 represent genomic DNA sequences used to illustrate the  
 CC method for determining the degree of cytosine methylation described in  
 CC the disclosure of the invention.  
 XX  
 SQ Sequence 512 BP; 100 A; 44 C; 145 G; 223 T; 0 other;

Query Match 9.5%; Score 33.6; DB 24; Length 512;  
 Best Local Similarity 63.8%; Pred. No. 3;  
 Matches 51; Conservative 0; Mismatches 29; Indels 0; Gaps 0;  
 QY 131 ATTTACGTTAATTGCAACAGATAACCAAGAAAGTGGACCAAGTCAAGTATTCGG 190  
 DB 100 AATTACGTTAATCGTAATAATAACCAATCACTAAATCTAAACAATTCATCATCGTT 41  
 QY 191 CAAACGGCTTACGGTAGCGG 210  
 DB 40 CGAACAACTACGTTTATCG 21

RESULT 14  
 ABO29519  
 ID ABO29519 standard; DNA; 512 BP.  
 XX  
 AC ABO29519;  
 XX  
 DT 12-JUL-2002 (first entry)  
 XX  
 DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 16110.  
 KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
 KW drug; side effect; cancer; central nervous system; cardiovascular;  
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
 KW SNP; cell differentiation; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200218632-A2.  
 XX  
 PD 07-MAR-2002.  
 XX  
 PF 01-SEP-2001; 2001WO-EP10074.  
 XX  
 PR 01-SEP-2000; 2000DE-1043826.  
 PR 05-SEP-2000; 2000DE-1044543.  
 XX  
 PA (EPIG-) EPIGENOMICS AG.  
 XX  
 PI Olek A, Piepenbrock C, Berlin K, Guetig D;  
 XX  
 DR WPI; 2002-371829/40.  
 XX  
 PT Determining the degree of cytosine methylation in genomic DNA, useful  
 PT for diagnosis and prognosis, comprises selective hybridization of  
 PT amplicons from chemically treated DNA  
 XX  
 PS Claim 12; 56pp + Sequence Listing; 56pp; German.

XX This invention describes a novel method for determining the degree of  
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
 CC genomic sample of DNA. The sample is treated chemically to convert  
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
 CC DNA that contains the target C is amplified to form a labeled amplicon.  
 CC The amplicon is hybridised to two classes, each with at least one  
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
 CC and the degree of hybridisation to both classes is determined from the  
 CC label on the amplicon. From the ratio of labels hybridised to the two  
 CC classes of oligomers, the degree of methylation is calculated. The method  
 CC is used: (i) for diagnosis and/or prognosis of side effects of

CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
CC systems etc., particularly by detecting mutations or single nucleotide  
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue  
CC types and for investigating cell differentiation. The method allows the  
CC methylation status of many C residues to be determined simultaneously.  
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the  
CC method for determining the degree of cytosine methylation described in  
CC the disclosure of the invention.  
XX  
SQ Sequence 512 BP; 223 A; 145 C; 44 G; 100 T; 0 other;

Query Match 9.5%; Score 33.6; DB 24; Length 512;  
Best Local Similarity 63.8%; Pred. No. 3;  
Matches 51; Conservative 0; Mismatches 29; Indels 0; Gaps 0;  
QY 131 ATTACGTTAAATGCAACAGAAATACCAAGAAAGCTGGACAGTCAAGTATTCGGG 190  
DB 413 AATTACGTTAAATGCAACAGAAATACCAAGAAAGCTGGACAGTCAAGTATTCGGG 190  
QY 191 CAACGGCTTACGTTAGCGC 210  
DB 473 CGAACAACTACGTTTAATCG 492

RESULT 15  
ABL70479/c  
ID ABL70479 standard; DNA; 9742 BP.  
XX  
AC ABL70479;  
XX  
DT 01-JUL-2002 (first entry)  
XX  
DE Chemically treated cell signalling DNA sequence#185.  
XX  
KW Cell signalling; cytosine methylation; cell signalling disease;  
XX cancer; tumour; cytostatic; ds.  
XX  
OS Unidentified.  
XX  
PN WO200202807-A2.  
XX  
PD 10-JAN-2002.  
XX  
PF 29-JUN-2001; 2001WO-EP07471.  
XX  
PR 30-JUN-2000; 2000DE-1032529.  
XX  
PR 01-SEP-2000; 2000DE-1043826.  
XX  
PA (EPIG-) EPIGENOMICS AG.  
XX  
PI Olek A, Piepenbrock C, Berlin K;  
XX  
DR WPI; 2002-154758/20.  
XX  
PT Nucleic acid, useful for diagnosis and therapy of diseases associated  
PT with cell signalling e.g. cancer, comprises chemically modified genomic  
PT sequences of genes associated with cell signalling  
PS Claim 1; SEQ ID NO 369; 24pp+sequence listing; English.  
XX

CC The invention relates to a nucleic acid comprising a sequence of at least  
CC 18 bases of a segment of chemically pretreated DNA of genes associated  
CC with cell signalling. The activity of the modified sequences of the  
CC invention may be described as cytostatic. The object of the invention is  
CC to provide the chemically modified DNA of genes associated with cell  
CC signalling, as well as oligonucleotides and/or PNA-oligonucleotides for  
CC detecting cytosine methylations, as well as a method which is  
CC particularly suitable for the diagnosis and/or therapy of genetic and  
CC epigenetic parameters of genes associated with cell signalling. The  
CC chemically modified DNA provided by the invention is useful for diagnosis  
CC and therapy of diseases such as solid tumours and cancer. The sequences  
CC given in records ABL70111-ABL70626 represent chemically pre-treated

CC genomic DNA's of genes associated with cell signalling.  
CC Note: The sequence data for this patent is not represented in the printed  
CC specification, but is based on sequence information supplied by the  
CC European Patent Office.

XX  
SQ Sequence 9742 BP; 3039 A; 88 C; 1939 G; 4676 T; 0 other;  
Query Match 9.5%; Score 33.6; DB 24; Length 9742;  
Best Local Similarity 52.9%; Pred. No. 8.3;  
Matches 72; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 51 TTATTGGAAGAAAAGGATTTACGTCCTATTCGGTATTCACAATACGGCTGATTACGGA 110  
DB 5177 TTTATATAAAAAAATAAATAATCAATAAATTCCTAAAAACGAAACCGCAAAATATTAT 5118  
QY 111 ATTCCTCAAAACGCTAAATAAATAAATAATCTACAAAATAACAAAAAACCCTT 5058  
DB 5117 AATCTCAAAACGCTAAATAAATAAATAATCTACAAAATAACAAAAAACCCTT 5058  
QY 171 GAACCACTCAAGTATT 186  
DB 5057 AATTAATCACTTAAT 5042

Search completed: November 15, 2003, 00:35:33  
Job time : 230.827 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 14, 2003, 23:58:39 ; Search time 2354.89 Seconds  
(without alignments)  
3643.257 Million cell updates/sec

Title: US-09-928-457-83

Perfect score: 353

Sequence: 1 CGGGAATTCGACGAGATG.....ACGGAGGACCGTTGGCTTT 353

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estmu:\*

5: em\_estov:\*

6: em\_estpl:\*

7: em\_estro:\*

8: em\_htc:\*

9: gb\_est1:\*

10: gb\_est2:\*

11: gb\_htc:\*

12: gb\_est3:\*

13: gb\_est4:\*

14: gb\_est5:\*

15: em\_estfun:\*

16: em\_estom:\*

17: em\_gss\_hum:\*

18: em\_gss\_inv:\*

19: em\_gss\_pln:\*

20: em\_gss\_vrt:\*

21: em\_gss\_fun:\*

22: em\_gss\_mam:\*

23: em\_gss\_mus:\*

24: em\_gss\_pro:\*

25: em\_gss\_rod:\*

26: em\_gss\_phg:\*

27: em\_gss\_vrl:\*

28: gb\_gss1:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	43.6	12.4	456	9	AA981355
C 2	39.8	11.3	498	28	AQ733611
C 3	39.6	11.2	1100	29	CNS014RI
C 4	39.6	11.2	1201	13	EX361080

5	39.4	11.2	853	29	CNS00CWP
6	38.6	10.9	771	14	CB998090
C 7	37.6	10.7	1101	29	CNS0032P
8	36.8	10.4	208	13	BU091084
C 9	36.8	10.4	1198	29	CC252073
C 10	36.8	10.4	1593	29	CC210958
11	36.6	10.4	967	14	CB562106
12	36.6	10.4	1018	13	EX400132
C 13	36.6	10.4	1101	29	CNS00ZCP
C 14	36.4	10.3	754	29	EX146252
15	36.2	10.3	499	28	EX146252
C 16	36.2	10.3	752	29	EX182757
C 17	36.2	10.3	1101	29	CNS0182P
C 18	36.2	10.3	1146	13	EX446722
C 19	36	10.2	551	29	EX244802
20	36	10.2	1201	13	EX375994
C 21	35.8	10.1	483	14	CB403766
C 22	35.8	10.1	1101	29	CNS00FEW
C 23	35.6	10.1	775	29	EX149557
C 24	35.4	10.0	331	12	BI315336
C 25	35.4	10.0	415	14	CD562547
C 26	35.4	10.0	464	10	EG019760
C 27	35.4	10.0	537	29	AG060614
C 28	35.4	10.0	550	12	BI447490
C 29	35.4	10.0	573	9	AW782350
C 30	35.4	10.0	772	14	CB201154
C 31	35.4	10.0	857	14	CB756442
C 32	35.4	10.0	866	29	BZ805512
C 33	35.4	10.0	949	29	CNS015RN
C 34	35.4	10.0	3909	11	AK085090
C 35	35.2	10.0	1201	9	AL514415
36	35	9.9	358	10	BE551716
C 37	34.8	9.9	557	28	BH336859
C 38	34.8	9.9	982	29	CNS006W7
C 39	34.8	9.9	1190	13	EX440012
40	34.8	9.9	1201	13	EX420578
41	34.6	9.8	601	29	EX207956
C 42	34.6	9.8	1200	13	EX414597
C 43	34.4	9.7	448	12	EG957424
44	34.4	9.7	544	13	BQ397837
45	34.4	9.7	567	28	AQ334267

## ALIGNMENTS

RESULT 1  
AA981355  
LOCUS  
DEFINITION  
vx56b03.r1 Stratagene mouse macrophage (H937306) Mus musculus cDNA  
Clone IMAGE:1279181.5, similar to TR:030357 030357 MODIFICATION  
METHYLASE ;, mRNA sequence.  
ACCESSION  
AA981355  
VERSION  
AA981355.1 GI:3159891  
KEYWORDS  
EST.  
SOURCE  
Mus musculus (house mouse)  
ORGANISM  
Mus musculus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 456)  
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
Waterston, R.  
The WashU-HMI Mouse EST Project  
Unpublished  
Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810

AA981355  
vx56b03.r1 Stratagene mouse macrophage (H937306) Mus musculus cDNA  
Clone IMAGE:1279181.5, similar to TR:030357 030357 MODIFICATION  
METHYLASE ;, mRNA sequence.  
ACCESSION  
AA981355  
VERSION  
AA981355.1 GI:3159891  
KEYWORDS  
EST.  
SOURCE  
Mus musculus (house mouse)  
ORGANISM  
Mus musculus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 456)  
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
Waterston, R.  
The WashU-HMI Mouse EST Project  
Unpublished  
Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810

Email: mouseest@watson.wustl.edu  
 This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium (infoimage.llnl.gov) for further information.  
 MG1:670981  
 Seq primer: -28ml3 rev1 ET from Amersham  
 High quality sequence stop: 440.  
 Location/Qualifiers

## FEATURES

source

1..456

/organism="Mus musculus"

/mol\_type="mRNA"

/db\_xref="taxon:10090"

/clone="IMAGE:1279181"

/tissue\_type="macrophage"

/dev\_stage="WEHI-3 cell line"

/lab\_host="SOLR (kanamycin resistant)"

/clone\_lib="Stratagene mouse macrophage (#937306)"

/note="Organ: blood; Vector: pBluescript SK-; Site 1:

EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer:

Oligo dt. WEHI-3 cell line. Average insert size: 1.5 kb;

Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTGGGACGAG

3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'

## BASE COUNT

ORIGIN

208 a

41 c

69 g

138 t

## Query Match

Best Local Similarity 12.4%; Score 43.6; DB 9; Length 456;

Matches 76; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY

37 ATAAATTCATAAGTTATTGGAAGAAAAAGGATTACCGTCGATTCGGTATTCACAATA 96

Db

242 AAAAAATATAATGACATATAAATATAGGATATCAGATAAATGTCACCACTTTAAATA 301

QY

97 CGGCTGATACCGAATTCGCCAAGCCGTAAAGATTACGTTAATTCGAACAGATAA 156

Db

302 GTGCAGATTATGAAGTTGTCACCAAGAAAAAGAGTAATTTTATAGCAATAAATA 361

QY

157 CCAAGAGAAA 166

Db

362 ACAAATTA 371

## RESULT 2

A0733611/c

LOCUS

HS\_5558\_A1\_C09\_T7A\_RPCI-11 Human Male BAC Library Homo sapiens

Genomic clone Plate=1134 Col=17 Row=E, genomic survey sequence.

ACCESSION

A0733611

VERSION

A0733611.1

KEYWORDS

GSS.

SOURCE

Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

Keller A., Shaker R., Furlong J., Young J., Zhao S., Adams M.D. and

Hood L.

TITLE

Sequence-tagged connectors: A sequence approach to mapping and

scanning the human genome

JOURNAL

Proc. Natl. Acad. Sci. U.S.A.

MEDLINE

99360589

PUBMED

10449764

COMMENT

Contact: Mahairas GG, Wallace JC, Hood L

High Throughput Sequencing Center

University of Washington

401 Queen Anne Avenue North, Seattle, WA 98109, USA

Tel: (206) 616-3618

Fax: (206) 616-3887

Email: jwallace@u.washington.edu

Clones are derived from the human BAC library RPCI-11. For BAC

or from Resear h Genetics (info@resgen.com). BAC end web Server:  
<http://www.husc.washington.edu>  
 Plate: 1134 row: E column: 17  
 Seq primer: T7  
 Class: BAC ends  
 High quality sequence stop: 498.  
 Location/Qualifiers

## FEATURES

source

1..498

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/clone="Plate=1134 Col=17 Row=E"

/sex="male"

/clone\_lib="RPCI-11 Human Male BAC Library"

/note="Vector: pBAC3.6; Site 1: EcoRI; Site 2: EcoRI;

Male blood DNA was isolated from one randomly chosen donor

and partially digested with a combination of EcoRI and

EcoRI Methylase. Size selected DNA was cloned into the

pBAC3.6 vector at EcoRI sites"

BASE COUNT 145 a

112 c

59 g

174 t

8 others

## Query Match

Best Local Similarity 11.3%; Score 39.8; DB 28; Length 498;

Matches 89; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY

14 CAGATGAAGAAAGCAGCGCTTGATAATTCATAAAGTTATTGGAAGAAAAAGGATTAC 73

Db

323 CAGTTTGAGAGATACACAGCGCTTTTGTGTGTCAGATCAGCATGAAGAGAGATTAG 264

QY

74 CTTCCATTTTCGTTATTCACAATACGCTGATTCGGAATTCGCCAAGCCGTAAGAATT 133

Db

263 CGTCATTGTATAAGTAAGATTAGGCTTAAGAAATGGAGCTTGGGGAACAAATATTT 204

QY

134 TACGTTAATTCGAACAGATTAACCAAGAAAGCTGGACCAAGTCAAGTA 184

Db

203 TACTTCAATTTGTGATGATCAATACCAACACCTCTCTAGAGAAATA 153

## RESULT 3

CNS014RI/c

LOCUS

Drosophila melanogaster

BACN12G24 of DrosBAC library from Drosophila melanogaster (fruit

fly), genomic survey sequence.

ACCESSION

AL104520

VERSION

AL104520.1

KEYWORDS

GSS.

SOURCE

Drosophila melanogaster (fruit fly)

ORGANISM

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

AUTHORS

Genoscope.

TITLE

Direct Submission

JOURNAL

Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :

BP 131 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr

- Web : www.genoscope.cns.fr)

COMMENT

Determination of this BAC-end sequence was carried out as part of a

collaboration with the European Drosophila Genome Project (EDGP) -

<http://www.edgp.ebi.ac.uk> - This Drosophila melanogaster BAC

library (DrosBAC) was made by Alain Billaud at CEPH (Centre

d'Etude du Polymorphisme Humain) with funding provided by a MRC

project grant. The DNA was prepared from embryos by Alain Bucheton

and Genevieve Payan. It has been constructed in the vector

pBelOBAC11.

FEATURES

Location/Qualifiers

1..1100

/organism="Drosophila melanogaster"

/mol\_type="genomic DNA"

/db\_xref="taxon:7227"

/clone="BACN12G24"



```

/clone_lib="DrosBAC"
/plasmid="pBelOBAC11"
/notes="end : S86"

BASE COUNT      184 a   175 c   118 g   190 t   433 others
ORIGIN

Query Match      11.2%; Score 39.6; DB 29; Length 1100;
Best Local Similarity 10.4%; Pred. No. 19;
Matches 26; Conservative 82; Mismatches 141; Indels 0; Gaps 0;

QY 1 CGGGAATCTGACGAGATGAAGAAGACGAGCTGATTAATTTCTAAGATTATGGAG 60
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 983 WGGGGGGMKGMMGMMGMMGMMGMMGMMGMMGMMGMMGMMGMMGMMGMMGMMGMM 924
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 61 AAAAAAGATTACCGTCCATTTCGGTATTCAATACGCGTCGATTACGGAATTCGCCAA 120
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 923 AAKKKKMMBMMCHMMKKKKMMKKMMKKMMKKMMKKMMKKMMKKMMKKMMKKMMKKMM 864
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 121 GCGTAAAGATTATCGTTTAATGTCACAAAGATAACCAAGAAAGCTGGAACGATCA 180
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 863 KMMMMKKMMNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 804
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 181 AGTATTCGGCAACGGCTTACGGTACGGATGTTTGGGAATGGAATGCTTCCCAA 240
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 803 KTNVANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 744
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 241 CATTATTC 249
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 743 MNNNNNTKN 735

RESULT 4
BX361080/c
LOCUS
DEFINITION
BX361080 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
clone CS0DI079YN16 3-PRIME, mRNA sequence.
ACCESSION
BX361080
VERSION
BX361080.1 GI:30374504
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
REFERENCE
1 (bases 1 to 1201)
AUTHORS
Li W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE
Full-length cDNA libraries and normalization
JOURNAL
Unpublished
COMMENT
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0DI079DG08NP1.
Location/Qualifiers
1..1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DI079YN16"
/tissue_type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-cligo (dt)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

BASE COUNT      155 a   152 c   130 g   182 t   582 others
ORIGIN

Query Match      11.2%; Score 39.6; DB 13; Length 1201;
Best Local Similarity 8.9%; Pred. No. 19;
Matches 22; Conservative 128; Mismatches 98; Indels 0; Gaps 0;

/clone_lib="DrosBAC"
/plasmid="pBelOBAC11"
/notes="end : S86"

BASE COUNT      184 a   175 c   118 g   190 t   433 others
ORIGIN

Query Match      11.2%; Score 39.6; DB 29; Length 1100;
Best Local Similarity 10.4%; Pred. No. 19;
Matches 26; Conservative 82; Mismatches 141; Indels 0; Gaps 0;

QY 1 CGGGAATCTGACGAGATGAAGAAGACGAGCTGATTAATTTCTAAGATTATGGAG 60
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 983 WGGGGGGMKGMMGMMGMMGMMGMMGMMGMMGMMGMMGMMGMMGMMGMMGMMGMM 924
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 61 AAAAAAGATTACCGTCCATTTCGGTATTCAATACGCGTCGATTACGGAATTCGCCAA 120
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 923 AAKKKKMMBMMCHMMKKKKMMKKMMKKMMKKMMKKMMKKMMKKMMKKMMKKMMKKMM 864
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 121 GCGTAAAGATTATCGTTTAATGTCACAAAGATAACCAAGAAAGCTGGAACGATCA 180
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 863 KMMMMKKMMNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 804
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 181 AGTATTCGGCAACGGCTTACGGTACGGATGTTTGGGAATGGAATGCTTCCCAA 240
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 803 KTNVANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 744
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 241 CATTATTC 249
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 743 MNNNNNTKN 735

RESULT 5
CNS00CWP
LOCUS
DEFINITION
Drosophila melanogaster genome survey sequence T7 end of BAC #
BACR26M06 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION
AL060036
VERSION
AL060036.1 GI:4947297
KEYWORDS
GSS.
SOURCE
Drosophila melanogaster (fruit fly)
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Spheroidea; Drosophilidae; Drosophila.
1 (bases 1 to 853)
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazuo Ogoe and
Aaron Mammeter in Pister de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
pi and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
Location/Qualifiers
1..853
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR26M06"
/clone_lib="RPCI-98"
/notes="end : T7"

BASE COUNT      280 a   162 c   158 g   187 t   66 others
ORIGIN

Query Match      11.2%; Score 39.4; DB 29; Length 853;
Best Local Similarity 48.3%; Pred. No. 21;
Matches 73; Conservative 11; Mismatches 67; Indels 0; Gaps 0;

37 ATAATTCATAAGATTATTGGAAGAAAAAGGATTACCGTCATTTCGGTATTCACAATA 96

```

	:   :	
461	A A A T T T T T A A A A A W W A A A A A A A A A A G A G A I T T W A T T T A A W C T G G G A A C G A A T A A A	520
Db		
97	C G G T G A T T C G G A A T T C C C A A G C G T A A A A G A G I T T A C G T T A A T T G C A A A C A G A A T A A	156
Qy		
521	A T A T T A A G T T A A A A A W W A A A A A A A A A A A A A A A T T A A A A A A A T A A T A A T A A A	580
Db		
157	C A A A G A A A A G C T G G A A C C A C T C A A G T A T T C	187
Qy		
581	A A A A A W A A A A T A M T A A A A A A A A A A A A A A A C A W T C	611
Db		

RESULT 6	CB998090	771 bp	mrNA	linear	EST 01-MAY-2003
LOCUS	CB998090				
DEFINITION	AGENCEOT 13890706 NIH MGC 148 Homo sapiens cDNA clone IMAGE:30378468 5', mrNA sequence.				
ACCESSION	CB998090				
VERSION	CB998090.1	GI:30292610			
KEYWORDS	EST.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 771)				
AUTHORS	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .				
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)				
JOURNAL	Unpublished				
COMMENT	Contact: Robert Strausberg, Ph.D.				

BASE COUNT	227 a	162 c	197 g	185 t	Library.
ORIGIN					
Query Match		10.9%;	Score 38.6;	DB 14;	Length 771;
Best Local Similarity		50.8%;	Pred. No. 32;		
Matches		92;	Conservative	0;	Mismatches 89; Indels 0; Gaps 0;
Qy	80	TTTCGGTATTCCACATACGCTGATTAACGGGAATTC	CCCCAAGCGCTAAAGAGATTTCAGTT	139	
Db	537	TGTCGGTGTTTACAAAATGGATTC	ATCATGACGACCTTCAGCCAGAGATATGAGA	596	
Qy	140	AATTGCCAACAGAAATACCAAAAGAAAAGCTGGAA	ACCAGTCAAGATATTCGGGCAAAACGGCT	199	
Db	597	AATTTGTTAAGAAATTCAGCACTTACATTAAGAAAAT	TGGCTACACCCCGACACAGTTTAC	656	

[illegible]

RESULT 7					
CNS0032P/c					
LOCUS	CNS0032P	1101 bp	DNA	linear	GSS 03-JUN-1999
DEFINITION	Drosophila melanogaster genome survey sequence T7 end of BAC # BAC07F18 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.				

```

BASE COUNT      330 a      188 c      431 t      119 others
ORIGIN

Query Match      10.7%   Score 37.6;   DB 29;   Length 1101;
Best Local Similarity 39.9%;   Freq. NO. 60;
Matches 87;   Conservative 26;   Mismatches 105;   Indels 0;   Gaps 0;

```

7	TTCTGACGAGTAAGAAAGACGAGCTTGATATATTTTCATATAAGTTATTTCGAAAGAAAAAG	66
15	TTAGGAACAGACAGAGAAATGAATGAAGATGAATCCCAAGGTTTATTCGAGAAGATT	74
67	GATTTACCGTCCATTCGGTATTCCATATACGCTGATTAGGAAATTCCTCCAAAGCCGTA	126
75	CAATTACATTCACATTTTCTCGGTCATATTTAAATGTTAAAAAATCACAATAATGATATA	134
127	AAAGATTTACGTTAATTCGAAACAGATAACCAAGAAAA	166
135	AAAAATTAATTTTATCTCTCAAAAAAATAAAAAA	174
CC252073	1198 bp DNA linear	GSS 13-MAY-2003
CH261-18M7	Sp6.1 CH261 Gallus gallus genomic clone	CH261-18M7,
genomic survey sequence.		
CC252073		
CC252073.1	GI:30598823	
GSS		
Gallus gallus (chicken)		
Gallus gallus		
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;		
Phasianinae; Gallus		
1 (bases 1 to 1198)		
Kramitzki, C., Higginbotham, J., Wylie, K., Carter, J., McPherson, J.,		
Warren, W., Graves, T., Mardis, E. and Wilson, R.		
Gallus gallus BAC End Reads		
Unpublished		
Contact: Richard K. Wilson		
Genome Sequencing Center		
Washington University School of Medicine		
Email: submissions@wustl.edu		
Insert Length: 182000 Std Error: 0.00		
Seq primer: Sp6 ATTTAGTGACACTATAG		
Class: BAC ends		
High quality sequence start: 32		
High quality sequence stop: 772.		
Location/Qualifiers		
1. 1198		
/organism="Gallus gallus"		
/mol_type="genomic DNA"		
/strain="Red Jungle Fowl"		
/db_xref="taxon:9031"		
/clone="CH261-18M7"		
/sex="female"		
/cell_line="UCD001, inbred 256"		
/clone_lib="CH261"		
/note="Vector: pTABBAC2.1; Site 1: EcoRI; Site 2: EcoRI;		
CH261 Female Chicken library - for library and clone		
ordering information: http://www.chori.org/bacpac		
BASE COUNT	365 a 202 c 271 g 355 t	
ORIGIN		
Query Match	10.4%; Score 36.8; DB 29; Length 1198;	
Best Local Similarity	50.6%; Pred. No. 96;	
Matches	89; Conservative 0; Mismatches 87; Indels 0; Gaps 0;	
86	TATTCAAAATACGGCTGATTACGGAATTCGCCAAAGCCGTAAGATTTACGTTAATTCG	145
357	TGTACAGATGATGATTTTTCGTCATCTTAAATACATGAATTTATTCATTCATTCG	298
146	AAACAGAAATACCAAGAAAGCTGGAACAGTCAGTTCAGTATTCGGCAACGGCTTACGGT	205
297	ATAACATTCGACACTGAAAGAGAAAGTCTCTTACCTGATCCTTCTTAATTACAGC	238
206	AGCCGATGTTTGGGAATGGAATGGCTTTCCCAACATTTATTCGAGGACACCAAGA	261
237	ATCTTCAGTTTCAGGAATGGAAGTCTCTACTATTAATTTGAATAAGACTCAT	182

```

RESULT 10
CC210958/c
LOCUS
DEFINITION
  CC210958      1593 bp      DNA      linear      GSS 12-MAY-2003
  CH261-170K18_RM1.1 CH261 Gallus gallus genomic clone CH261-170K18,
  genomic survey sequence.
ACCESSION
  CC210958
VERSION
  CC210958.1 GI:30523626
KEYWORDS
  GSS.
SOURCE
  Gallus gallus (chicken)
  Gallus gallus
  ORGANISM
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
    Phasianinae; Gallus.
REFERENCE
  1 (bases 1 to 1593)
  Krenitzki,C., Higginbotham,J., Wylie,K., Carter,J., McPherson,J.,
  Warren,W., Graves,T., Mardis,E. and Wilson,R.
  Gallus gallus BAC End Reads
  Unpublished
  Contact: Richard K. Wilson
  Genome Sequencing Center
  Washington University School of Medicine
  Email: submissions@wustl.edu
  Insert Length: 182000 Std Error: 0.00
  Seq primer: RM1 TACGACTCACTATAGGAGA
  Class: BAC ends
  High quality sequence start: 360
  High quality sequence stop: 1045.
  Location/Qualifiers
    1..1593
      /organism="Gallus gallus"
      /mol_type="genomic DNA"
      /strain="Red Jungle Fowl"
      /db_xref="taxon:9031"
      /clone="CH261-170K18"
      /sex="female"
      /cell_line="UCD001, inbred 256"
      /clone_lib="CH261"
      /note="Vector: pTAC2.1; Site 1: EcoRI; Site 2: EcoRI;
      CH261 Female Chicken library - For library and clone
      ordering information: http://www.chori.org/bacpac"
BASE COUNT      390 a      219 c      304 g      385 t      295 others
ORIGIN
  Query Match      10.4%; Score 36.8; DB 29; Length 1593;
  Best Local Similarity 50.6%; Pred. No. 1e+02;
  Matches 89; Conservative 0; Mismatches 87; Indels 0; Gaps 0;
  QY      86 TATTCAATACGGTGTATTAGGAAATCCCGAAGCGTAAAGATTAGCTTAATGTC 145
  Db      632 TGTACACAGTATGATTTTGTGTCATCTTAAATACATTTGAATATTATTATTCAATGG 573
  QY      146 AATACAGTAATACCAAGAAAGCTGGAACAGTATTCGGCAACGCTTACGGT 205
  Db      572 ATAAACATTGCACCTGAAGAAAGAGAAAGTCTCTAACCTGATCTTTCTTAATACAGC 513
  QY      206 AGCCGATGTTTGGGAATGGAATGGCTTCCCAATATTATTCAGGACACCAAGA 261
  Db      512 AICTTCAGTTCAGGAATGGAAGTGTCTACTAATATTATGTAATAGACTCATA 457

RESULT 11
CB562106
LOCUS
DEFINITION
  AGENCOURT I3321331 NICHD XGC Tad1 Xenopus laevis cDNA clone
  IMAGE:6675382 3', mRNA sequence.
ACCESSION
  CB562106
VERSION
  CB562106.1 GI:29481636
KEYWORDS
  EST.
SOURCE
  Xenopus laevis (African clawed frog)
  Xenopus laevis
  ORGANISM
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae;
    Xenopodinae; Xenopus.

```

```

REFERENCE
  1 (bases 1 to 967)
  NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
  Tumor Gene Index
  Unpublished
  Contact: Robert Strausberg, Ph.D.
  Email: cgapbs-r@mail.nih.gov
  Tissue Procurement: Drs. Donald Brown and Liquean Cai
  cDNA Library Preparation: CLONTECH
  cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
  DNA Sequencing by: Agencourt Bioscience Corporation
  Clone distribution: NCI-CGAP clone distribution information can be
  found through the I.M.A.G.E. Consortium/LLNL at:
  http://image.llnl.gov
  Plate: LNC9311 row: b column: 21
  High quality sequence start: 15
  High quality sequence stop: 517.
  Location/Qualifiers
    1..967
      /organism="Xenopus laevis"
      /mol_type="mRNA"
      /db_xref="taxon:8355"
      /clone="IMAGE:6879382"
      /dev_stage="metamorphosis stage 53"
      /clone_lib="NICHD XGC Tad1"
      /note="Organ: Developing Tadpole; Vector: pDNR-LIB;
      Site 1: Sfi; Site 2: Sfi; 5' and 3' adaptors were used in
      cloning as follows: 5' adaptor sequence:
      5'-CACGCCATTATGGCC-3' and 3' adaptor sequence:
      5'-ATTTAGAGCGGCGCCGACATG-df(30)BN-3' (where B = A,
      C, or G and N = A, C, G, or T). Average insert size 1.6 kb
      (range 0.9-3.0 kb). 15/15 colonies contained inserts by
      PCR. This library was enriched for full-length clones and
      was constructed by Clontech Laboratories (Palo Alto, CA)."
BASE COUNT      347 a      166 c      186 g      268 t
ORIGIN
  Query Match      10.4%; Score 36.6; DB 14; Length 967;
  Best Local Similarity 56.1%; Pred. No. 1e+02;
  Matches 69; Conservative 0; Mismatches 54; Indels 0; Gaps 0;
  QY      51 TTATTCGACAAAGCGTAAAGATTACCGTCCATTTCGGTATTTCACATACGCTGATTACGGA 110
  Db      499 TTATTTCAATAAAGGAACCTTAACCCCTTTTACTGCGAAAGATTGGCTGTATCTTT 558
  QY      111 ATTCCTCCAAAGCGTAAAGATTACGTTAATTTCGCAACAGAAATACCAAGAAAGCTG 170
  Db      559 GAGAGACACAGGCTTAAGTAAGTACCTAGCTATATATAAAATAAAACCAAAAAAATG 618
  QY      171 GAA 173
  Db      619 GAA 621

RESULT 12
BX400132
LOCUS
DEFINITION
  BX400132 Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED
  Homo sapiens cDNA clone CS0DJ004Y23 5-PRIME, mRNA sequence.
ACCESSION
  BX400132
VERSION
  BX400132.1 GI:30618078
KEYWORDS
  EST.
SOURCE
  Homo sapiens (human)
  Homo sapiens
  ORGANISM
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 1018)
  Li,W.B., Gruber,C., Jesse,J. and Polayes,D.
  Full-length cDNA libraries and normalization
  Unpublished
  Contact: Genoscope
  Genoscope - Centre National de Sequencage
  BP 191 91006 EVRY cedex - France

```

Email: [segref@genoscope.cns.fr](mailto:segref@genoscope.cns.fr), Web: [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 5199.f For  
more information about this cluster, see  
<http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0DJ004YM23>  
Feng Liang Email: [liang@lifetech.com](mailto:liang@lifetech.com) URL:  
<http://fulllength.invitrogen.com/> Invitrogen Corporation 1600  
Paradise Avenue Genoscope sequence ID: CS0DJ004AG12QP1.  
Location/Qualifiers

FEATURES  
source

```
1..1018
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DJ004YM23"
/cell_type="T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"
/cell_line="JURKAT"
/clone_lib="Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT      291 a 116 c 175 g 238 t 198 others
ORIGIN
```

```
Query Match      10.4%; Score 36.6; DB 13; Length 1018;
Best Local Similarity 27.0%; Pred. No. 1.1e-02;
Matches 80; Conservative 75; Mismatches 14; Indels 0; Gaps 0;

QY 19 TGAAGAAGCAGCGTTCGATATTCATAAGTTATTGGAAGAAAGGATTTACCGTCC 78
Db 717 TTWTAAGAAAGAAAGRVRTTTTAAARVGGGAGGGAAGAAWATAATGGGK 776
QY 79 ATTTCGGTATTCACATACGGGTGATTACGGAATTCCTCCAAAGCGTAAAGATTTACGT 138
Db 777 TTTTGTGAGGAAGAAAGTGGGGGGGGAATTTGGRAAAMVACVHMGGGTTTAA 836
QY 139 TAATTGCAACAGATATACCAAGAAAGCTGGACCACTCAAGTATTCGGCGCAACGGC 198
Db 837 MMAGGGAAGAAAGAAWTTTWWGGKTAAKKTAAAMARVKGDKDTWAKYDKKKK 896
QY 159 TTACGGTAGCCGATGTTTGGGAATGGAATGGGTTTCCCAACATTTATTCAGACACCA 258
Db 897 TTKMMWAKKDKKKKDDTAAAGAAATTTATTATAAAATTTTAAAKWCM 956
QY 259 AGACGAACGATTTTATGCTAGCTGTGCGGGAATATCTGATATCACTTGAACG 314
Db 957 MTCMMKVCANVMCMVMCMVMCMVMCMVMCMVMCMVMCMVMCMVMCMVMCMVMCM 1012
```

```
RESULT 13
LOCUS      CNS002CP/c
DEFINITION Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN01312 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION  AL097507
VERSION     AL097507.1 GI:5609118
KEYWORDS   GSS.
SOURCE     Drosophila melanogaster (fruit fly)
ORGANISM   Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
            1 (bases 1 to 1101)
REFERENCE   Genoscope.
            Direct Submission
            Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr)
            - Web : www.genoscope.cns.fr
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
```

<http://www.edgp.ebi.ac.uk> - This Drosophila melanogaster BAC  
library (Dros BAC) was made by Alain Billaud at CEPH (Centre  
d'Etude du Polymorphisme Humain) with funding provided by a MRC  
project grant. The DNA was prepared from embryos by Alain Bucheton  
and Genevieve Payan. It has been constructed in the vector  
pBeloBAC11.

FEATURES  
source

```
1..1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACN01E12"
/clone_lib="DrosBAC"
/plasmid="pBeloBAC11"
/notes="end : SP6"
BASE COUNT      346 a 170 c 129 g 306 t 150 others
ORIGIN
```

```
Query Match      10.4%; Score 36.6; DB 29; Length 1101;
Best Local Similarity 39.3%; Pred. No. 1.1e-02;
Matches 64; Conservative 28; Mismatches 71; Indels 0; Gaps 0;

QY 3 GGAATTCGAGCAGATGAAGAAAGCAGCGCTTGATATTCATAAGTATTGGAAGAA 62
Db 1090 KDRWAAAKWETRWAAAAAFAAAAAAADTTTTTTWAAARAWRWTTTTTTTAAAA 1031
QY 63 AAAGGATTTACCGTCCATTTCCGTATTTCACATACGCTGATTACGGAATTCCTCC 122
Db 1030 AAAATTTTWAAGTTTTTKKTTTAAATTAATAAAAAATTTTDAWAAAAATTTTAA 971
QY 123 CGTAAAGATTTACGTTAATTCGAAACAGATAACCAAGAA 165
Db 970 TAAAAWAAAWTTGGTTTAAWAAAAAATAAAAAATTTAAAA 928
```

```
RESULT 14
LOCUS      BX146252/c
DEFINITION Danio rerio genomic clone DKEY-10312, genomic survey sequence.
ACCESSION  BX146252
VERSION     BX146252.1 GI:27977610
KEYWORDS   GSS.
SOURCE     Danio rerio (zebrafish)
ORGANISM   Danio rerio
```

```
REFERENCE     1 (bases 1 to 754)
AUTHORS       Humphray, S.J., Huckle, E. and Durham, J.L.
TITLE         Direct Submission
JOURNAL       Submitted (13-MAR-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquerry@sanger.ac.uk Unpublished
            This sequence was generated from the T7 end of BAC 10312. 10312 is
            part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D\_rerio/.
```

FEATURES  
source

```
1..754
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="DKEY-10312"
/tissue_type="testis"
/notes="vector pIndigoBAC-536"
BASE COUNT      234 a 156 c 129 g 235 t
ORIGIN
```

```
Query Match      10.3%; Score 36.4; DB 29; Length 754;
Best Local Similarity 51.9%; Pred. No. 1.1e-02;
Matches 82; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 105 TAGGAATTCCTCCAAAGCGTAAAGATTTAGCTTAATTCGAAACAGATAACCAAGAA 164
```

Db 747 TACTGCATATCCAAATATCTTACAAATAGTAGTATCCATGTACCAAAACAAACAGAG 688  
 QY 165 AAGCTGAACAGTCFAAGTATTCGGGCAACGGCTTACGGTAGCCGATGTTTGGGAATG 224  
 Db 687 AAGCTGGAGAAAGAGGAGCTATCTGGAATATCTGATCATGTATATATACTTTTAGGAGCT 628  
 QY 225 GAAATGCTTTCCCAACATTTTCAGAGACACCAAGAC 262  
 Db 627 GAAAAGAGCGACACCATGTTAGGGCTAGATTCTGAGCC 590

RESULT 15  
 BH030890  
 LOCUS  
 DEFINITION  
 BH030890  
 ACCESSION  
 VERSION  
 BH030890.1 GI:14800053  
 KEYWORDS  
 GSS  
 SOURCE  
 Mus musculus (house mouse)  
 ORGANISM  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 499)  
 Zhao S., Nierman W., Malek J., Shatsman S., Akinret B., Levins M.,  
 Tsegaye G., Geer K., Krol M., Shvartsbeyn A., Gebregeorgis E.,  
 Russell D., de Jong P. and Fraser C.M.  
 Mouse BAC End Sequences from Library RPCI-24  
 UNPUBLISHED  
 Other GSSs: RPCI-24-23914.TV  
 Contact: Shaying Zhao  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: szhao@tigr.org  
 Clones are derived from the mouse BAC library RPCI-24. For BAC  
 library availability, please contact Pieter de Jong  
 (pdejong@mail.cho.org). Clones may be purchased from BACPAC  
 Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end  
 page: [http://www.tigr.org/tdb/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html)  
 Plate: 239 row: I column: 4  
 Seq primer: SP6  
 Class: BAC ends.

FEATURES  
 source  
 1..499  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="RPCI-24-23914"  
 /sex="Male"  
 /cell\_type="Spleen/Brain"  
 /clone\_lib="RPCI-24"  
 /note="Vector: pTARBAC1; Site 1: BamHI; Site 2: BamHI;  
 RPCI-24 Mouse BAC Library produced by Pieter de Jong. The  
 library was cloned in the pTARBAC1 cloning vector at the  
 BamHI sites using MboI partially digested male C57BL/6J  
 DNA."  
 DNA. 74 c 114 g 163 t

BASE COUNT 148 a 74 c 114 g 163 t  
 ORIGIN  
 Query Match 10.3%; Score 36.2; DB 28; Length 499;  
 Best Local Similarity 50.3%; Pred. No. 1.2e+02;  
 Matches 89; Conservative 0; Mismatches 88; Indels 0; Gaps 0;  
 QY 15 AGAATCAAGAGAGCGGCTTCATATTTTCAATAGTTATTGGAGAAAAAGGATTACC 74  
 Db 65 AGAGTGAGAGAGAGTGTTGTGATGTGAGAAATAGTCATTTCCAAAAAGAAACATGGT 124  
 QY 75 GTCCATTTCCGTTATTCACAAATACGGCTGATTACGGATTCCCAAGGCCGTAAGAATTT 134

Db 125 GAGTATTTTGTCTATTCACTTACCAGGTGATTGGGTATTTACTGTAAATATAAACAATAGG 184  
 QY 135 ACGTTAATTGCAACAGAGATTAACCAAGAAAGCTGGAAACCAAGTCAAGTATTTCGGGC 191  
 Db 185 ATTTAAATAGTGCCTAGCACATAGACAAATACCATGTTAGCTGTTATTTTAAAGTC 241

Search completed: November 15, 2003, 08:02:35  
 Job time : 2358.89 secs



```
Db 979 AAGAAGGCCATTTTCAGGAGTTTATTAAAGCTACTTAAAGAGTTAGAGTATAACTATATA 920
QY 81 TTCGGTATTCACATACCGGTGATTCAGGAATTCGCCAAAGCCCTAAAGATTACGTTA 140
Db 919 TCTTTTATAGCAATGCTGAGAACTATGGGATTCGCCAAAGAGAAAGAAAGACTCGTGCTC 860
QY 141 ATTGCAACAGAAATAACCAAGAAAGCT 169
Db 859 TTAGTAGTCGAGTAGGTAAGATTACCT 831

RESULT 2
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZ9pt-Fls
US-08-232-463-14

Query Match 10.2%; Score 36; DB 1; Length 7218;
Best Local Similarity 1.7%; Pred. No. 0.17;
Matches 6; Conservative 197; Mismatches 147; Indels 0; Gaps 0;

QY 2 GGAATTCGACAGCAATGAAGAAGCAGGCTTCATTAATTCATAAGTTATTGGAAGA 61
Db 1406 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1347
QY 62 AAGAAGATTACCGTCCATTTTCGGTATTCAACAATACGCTGATTACGGAATTCGCCAAG 121
Db 1346 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1287
QY 122 CGGTAAGATTACGTTAATTGCAACAGATAACCAAGAAAGCTGGAAACCAAGTCAA 181
```

```
Db 1286 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1227
QY 182 GTATTCGGGCAACGGCTTACGTTAGCCGATGTTTGGGAATGGAATGGCTTTCCCAAC 241
Db 1226 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1167
QY 242 ATTATTGCAAGACACCAAGACGAAACGGATTATTATGCATAGCTGTCCGGGAATTATCTGA 301
Db 1166 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1107
QY 302 TATCACTTGAACGATGGCTTGATACCTAAACACGAGAACCGTTGGCT 351
Db 1106 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1057

RESULT 3
US-09-328-111-626/c
; Sequence 626, Application US/09328111
; Patent No. 6262333
; GENERAL INFORMATION:
; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carroll III, Eddie
; APPLICANT: Catino, Theodore J.
; APPLICANT: Derti, Adrian
; APPLICANT: Ford, Donna M.
; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.
; APPLICANT: Schlegel, Robert
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/328,111
; CURRENT FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: US 60/098,801
; EARLIER FILING DATE: 1998-06-10
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 626
; LENGTH: 935
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)..(935)
; OTHER INFORMATION: n = A,T,C or G
US-09-328-111-626

Query Match 10.0%; Score 35.4; DB 3; Length 935;
Best Local Similarity 60.0%; Pred. No. 0.12; Indels 0; Gaps 0;
Matches 57; Conservative 0; Mismatches 38;

QY 129 AGATTTTACGTTAATTGCAACAGATAACCAAGAAAGCTGGAACCAAGCTTAATTGACCAATTTTACG 188
Db 342 AATTTTTCACCCCTCCCAAGGGAANCAAGGCTTAATTGACCAATTTTACG 283
QY 189 GCAAAACGGCTTACGTTAGCCGATGTTTGGGAAT 223
Db 282 GCCAAAGGCTTTGGGTACCCGAAAGCTTCGGGATT 248

RESULT 4
US-09-453-702B-158
; Sequence 158, Application US/09453702B
; Patent No. 6365723
; GENERAL INFORMATION:
; APPLICANT: Blattner, Frederick R.
; Burland, Valerie
; Perna, Nicole T.
```



```

; Plunkett, Guy
; Welch, Rod
; TITLE OF INVENTION: No. 6345723e1 Sequences of E. coli O157
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 8.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/453,702B
; FILING DATE: 03-Dec-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/110,955
; FILING DATE: 04-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27386
; REFERENCE/DOCKET NUMBER: 960296,95017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 251-5000
; TELEFAX: (608) 251-9166
; INFORMATION FOR SEQ ID NO: 158:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17136
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 158:
US-09-453-702B-158
Query Match 9.5%; Score 33.6; DB 4; Length 17136;
Best Local Similarity 57.7%; Pred. No. 1.3;
Matches 60; Conservative 0; Mismatches 44; Indels 0; Gaps 0;
QY 42 TTCATAAGTATTGCAAGAAAGGATTACCGTCCATTTCGGTATTTCACAAATACGGCT 101
Db 13604 TTCCTCGTATTGCCAGCAAGGATATACATACAAAACGATTTCTTCT 13663
QY 102 GATTACGGAATCCCAAGCGGTAAGAAAGATTACGTTAATGC 145
Db 13664 CATGATAAAATTTCCACGGCGTTATGATGATTGATTGATTC 13707

; Plunkett, Guy
; Welch, Rod
; TITLE OF INVENTION: No. 6345723e1 Sequences of E. coli O157
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 8.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/453,702B
; FILING DATE: 03-Dec-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/110,955
; FILING DATE: 04-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27386
; REFERENCE/DOCKET NUMBER: 960296,95017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 251-5000
; TELEFAX: (608) 251-9166
; INFORMATION FOR SEQ ID NO: 158:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17136
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 158:
US-09-453-702B-158
Query Match 9.5%; Score 33.6; DB 4; Length 17136;
Best Local Similarity 57.7%; Pred. No. 1.3;
Matches 60; Conservative 0; Mismatches 44; Indels 0; Gaps 0;
QY 42 TTCATAAGTATTGCAAGAAAGGATTACCGTCCATTTCGGTATTTCACAAATACGGCT 101
Db 13604 TTCCTCGTATTGCCAGCAAGGATATACATACAAAACGATTTCTTCT 13663
QY 102 GATTACGGAATCCCAAGCGGTAAGAAAGATTACGTTAATGC 145
Db 13664 CATGATAAAATTTCCACGGCGTTATGATGATTGATTGATTC 13707

; Plunkett, Guy
; Welch, Rod
; TITLE OF INVENTION: No. 6345723e1 Sequences of E. coli O157
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 8.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/453,702B
; FILING DATE: 03-Dec-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/110,955
; FILING DATE: 04-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27386
; REFERENCE/DOCKET NUMBER: 960296,95017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 251-5000
; TELEFAX: (608) 251-9166
; INFORMATION FOR SEQ ID NO: 158:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17136
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 158:
US-09-453-702B-158
Query Match 9.5%; Score 33.6; DB 4; Length 17136;
Best Local Similarity 57.7%; Pred. No. 1.3;
Matches 60; Conservative 0; Mismatches 44; Indels 0; Gaps 0;
QY 42 TTCATAAGTATTGCAAGAAAGGATTACCGTCCATTTCGGTATTTCACAAATACGGCT 101
Db 13604 TTCCTCGTATTGCCAGCAAGGATATACATACAAAACGATTTCTTCT 13663
QY 102 GATTACGGAATCCCAAGCGGTAAGAAAGATTACGTTAATGC 145
Db 13664 CATGATAAAATTTCCACGGCGTTATGATGATTGATTGATTC 13707

```

```

; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/936,165A
; FILING DATE: 24-SEP-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/027,032
; FILING DATE: 24-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Gimmi, Edward R.
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: P50549
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 960 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; US-08-936-165A-46
Query Match 9.3%; Score 33; DB 4; Length 960;
Best Local Similarity 51.7%; Pred. No. 0.65;
Matches 75; Conservative 0; Mismatches 70; Indels 0; Gaps 0;
QY 21 AAGAAGAGAGGCTTCATAATTTTCATAAGTATTTCGAGAGAAAGGATTACCGTCCAT 80
Db 882 AAGAAGAGTAAATGATTGTTTTTTCGAGAGAGATAAATGAATCATCGGAATTCCTA 823
QY 81 TTCGGTATTACAAATACGGCTGATTACGGAATTTCCCAAGCGTAAAGATTAGCGTTA 140
Db 822 TTCTGTATGCCITGATGTAATGCTAAATCGTAAATGTAATGTTAAAGTAAATTAATT 763
QY 141 ATTGCAACAGATAACCAAGAAA 165
Db 762 ATAGAAAACATAAATTCACAAAA 738

RESULT 6
US-09-385-982-505
; Sequence 505; Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CDNA-280XX
; CURRENT APPLICATION NUMBER: US/09/385,982
; CURRENT FILING DATE: 1999-08-30
; EARLIER APPLICATION NUMBER: 09/328,111
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: 60/117,393
; EARLIER FILING DATE: 1999-01-27
; EARLIER APPLICATION NUMBER: 60/098,639
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 505
; LENGTH: 628
; TYPE: DNA

```

ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(528)  
OTHER INFORMATION: n = A,T,C or G  
US-09-385-982-505

Query Match 9.2%; Score 32.4; DB 3; Length 628;  
Best Local Similarity 52.8%; Pred. No. 0.85;  
Matches 66; Conservative 0; Mismatches 59; Indels 0; Gaps 0;  
QY 147 AACGAAATACCAAGAAAGCTGGAACAGCTCAAGTATTCGGGCAACGGCTTACGGTA 206  
DB 448 AAAAAAGAGGAAATTAAGTTTATTCCTTTTGGGTTGAAGNATGAA 507  
QY 207 GCCGATGTTTGGGAATGGAATGGCTTCCCAACATTATTCAGGACACCAAGACGAAA 266  
DB 508 GGGGAAAATTTGGGCCITGAAATGTTTAAGTAATACTTTTCTAATAAGGTAAGTCTAGA 567  
QY 267 CGGAT 271  
DB 568 AGAAT 572

RESULT 7  
US-08-545-528D-1/c  
Sequence 1, Application US/08545528D  
Patent No. 6537773  
GENERAL INFORMATION:  
APPLICANT: Fraser et al.  
TITLE OF INVENTION: Nucleotide Sequence of the Mycoplasma Genitalium Genome, Fragment  
Patent No. 6537773  
TITLE OF INVENTION: Thereof, and Uses Thereof  
FILE REFERENCE: PB193P1  
CURRENT APPLICATION NUMBER: US/08/545.528D  
PRIOR FILING DATE: 1995-10-19  
PRIOR APPLICATION NUMBER: US 08/488,018  
PRIOR FILING DATE: 1995-06-07  
PRIOR APPLICATION NUMBER: US 08/473,545  
PRIOR FILING DATE: 1995-06-07  
NUMBER OF SEQ ID NOS: 1  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 1  
LENGTH: 580073  
TYPE: DNA  
ORGANISM: Mycoplasma genitalium  
US-08-545-528D-1

Query Match 9.1%; Score 32; DB 4; Length 580073;  
Best Local Similarity 54.2%; Pred. No. 16;  
Matches 65; Conservative 0; Mismatches 55; Indels 0; Gaps 0;  
QY 42 TCATAAAGTTATTGCAAGAAAGGATTTACCGTCCATTTCGGTATTCACAAATACGGCT 101  
DB 505245 TTAAAGAGTTATCTAAACAAAAGAAATTTACTGAGGATTTAGAAAGTAAACTAAACC 505186  
QY 102 GATTACGGAATTCGCCAAGCGGTAAAGATTACGTTAATGCAACAGAAATACCAAA 161  
DB 505185 GCTTTTAAATGCTTGTGAAACGTTTATCATGATTAACAGATTATGATATTACCAA 505126

RESULT 8  
US-09-107-532A-2880  
Sequence 2880, Application US/09107532A  
Patent No. 6583275  
GENERAL INFORMATION:  
APPLICANT: Lynn A Doucette-Stamm and David Bush  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
NUMBER OF SEQUENCES: 7310  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
STREET: 100 Beaver Street

CITY: Waltham  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02354  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: PC  
OPERATING SYSTEM: <Unknown>  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/107,532A  
FILING DATE: 30-Jun-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/085,598  
FILING DATE: 14 May 1998  
APPLICATION NUMBER: 60/051571  
FILING DATE: July 2, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Atinello, Pamela Deneka  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: GTC-012  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-5007  
TELEFAX: (781)893-8277  
INFORMATION FOR SEQ ID NO: 2880:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1353 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Enterococcus faecium  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (B) LOCATION 1...1353  
SEQUENCE DESCRIPTION: SEQ ID NO: 2880:  
US-09-107-532A-2880

Query Match 9.0%; Score 31.6; DB 4; Length 1353;  
Best Local Similarity 54.5%; Pred. No. 2;  
Matches 85; Conservative 0; Mismatches 59; Indels 2; Gaps 1;  
QY 126 AAAAGATTACGTTAATTGCAACAGAAATACCAAGAAAGCTGGAACCTCAAGTAT 185  
DB 470 AAAATTTATGGCGTTGTTTAGAAAAATATACCAAAAAAGAGAGAAATTACGCAAGGAT 529  
QY 186 TCGGCAACGGCTTACGGTAGCCGATGTTTGGGAATGGAATGCTTCCCAACATTA 245  
DB 530 TTGCGCCAGATGCTGA--ATTTCCTTTAATTTATCGAAAAAAGGCTTATTGCAAGCTTA 587  
QY 246 TTGCAGGACCAACAGCAACGGAATTTATGCATA 281  
DB 588 CTTCAGGACCCAGGACGACGAGTTTCGTAAA 623

RESULT 9  
US-08-916-421B-1/c  
Sequence 1, Application US/08916421B  
Patent No. 6503729  
GENERAL INFORMATION:  
APPLICANT: Bult et al.  
TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococcus  
PATENT NO. 6503729  
TITLE OF INVENTION: jannaschii  
FILE REFERENCE: PB275  
CURRENT APPLICATION NUMBER: US/08/916,421B  
CURRENT FILING DATE: 1997-08-22  
PRIOR APPLICATION NUMBER: US 60/024,428  
PRIOR FILING DATE: 1996-08-22  
NUMBER OF SEQ ID NOS: 3

```
SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1664976
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (28222)..(28222)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (28257)..(28258)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84773)..(84773)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84808)..(84808)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84812)..(84812)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98120)..(98120)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98159)..(98159)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98239)..(98239)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98266)..(98266)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98343)..(98343)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (103958)..(103958)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (148948)..(148948)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (163385)..(163385)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (191989)..(191989)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (191995)..(191995)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (231980)..(231980)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234187)..(234187)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (779676)..(779676)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (855539)..(855539)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (871619)..(871619)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1084830)..(1084830)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1096846)..(1096846)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1119881)..(1119881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1130881)..(1130881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1310988)..(1310988)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1313224)..(1313224)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1349473)..(1349473)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1349491)..(1349491)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1470091)..(1470091)
```

OTHER INFORMATION: n equals a, t, c, or g  
 NAME/KEY: misc feature  
 LOCATION: (1569020)..(1569020)  
 OTHER INFORMATION: n equals a, t, c, or g  
 NAME/KEY: misc feature  
 LOCATION: (1602912)..(1602912)  
 OTHER INFORMATION: n equals a, t, c, or g  
 NAME/KEY: misc feature  
 LOCATION: (1603734)..(1603734)  
 OTHER INFORMATION: n equals a, t, c, or g  
 NAME/KEY: misc feature  
 LOCATION: (1637998)..(1637998)  
 OTHER INFORMATION: n equals a, t, c, or g  
 NAME/KEY: misc feature  
 LOCATION: (1664855)  
 OTHER INFORMATION: n equals a, t, c, or g  
 US-08-916-421B-1

Query Match 9.0%; Score 31.6; DB 4; Length 1664976;  
 Best Local Similarity 50.7%; Pred. No. 31;  
 Matches 76; Conservative 0; Mismatches 74; Indels 0; Gaps 0;  
 QY 17 AATGAAGAAGACGAGCTTGATATTCATTAAGTTTGGAGAGAAAAGGATTTACCGT 76  
 DB 104475 AAGGAAGAAGAGATATTTTAAATAATTGAAGGAATTGAATTAAGGAGGATTTGATA 104416  
 QY 77 CCATTTCGGTATTCACATACGCTGATTACGGATTCGCCAAGCCGTAAAGATTTAC 136  
 DB 104415 AATATATCTTATTTCTCCAGATTAATCAATCAATAGGAACGGAATTTGTTG 104356  
 QY 137 GTTAATTCACACAGAAATACCAAGAAAA 166  
 DB 104355 AAAAACTTAACCAACAAATTCAGATGAGA 104326

RESULT 10  
 US-08-485-158A-6/c  
 ; Sequence 6, Application US/08485158A  
 ; Patent No. 5859328  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Nasrallah, June B.  
 ; APPLICANT: Nasrallah, Mikhail E.  
 ; APPLICANT: Thorness, Mary K.  
 ; TITLE OF INVENTION: ISOLATED DNA ELEMENTS THAT DIRECT  
 ; TITLE OF INVENTION: PISTIL-SPECIFIC AND ANOTHER-SPECIFIC GENE EXPRESSION  
 ; TITLE OF INVENTION: AND METHODS OF USING SAME  
 ; NUMBER OF SEQUENCES: 9  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Sughrue, Mion, Zinn, Macpeak, & Seas  
 ; STREET: 2100 Pennsylvania Avenue  
 ; CITY: Washington  
 ; STATE: D.C.  
 ; COUNTRY: U.S.A.  
 ; ZIP: 20037-3202  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/485.158A  
 ; FILING DATE: 07-JUN-1995  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; NAME: Mack, Susan J.  
 ; REGISTRATION NUMBER: 30,951  
 ; REFERENCE/DOCKET NUMBER: A-6217-1  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (202) 293-7060  
 ; TELEFAX: (202) 293-7860  
 ; TELEX: 6491103  
 ; INFORMATION FOR SEQ ID NO: 6:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 475 base pairs

TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 US-08-485-158A-6  
 Query Match 8.8%; Score 31; DB 2; Length 475;  
 Best Local Similarity 52.8%; Pred. No. 2.1;  
 Matches 67; Conservative 0; Mismatches 60; Indels 0; Gaps 0;  
 QY 39 AATTTCATAAAGTTATTGGAAGAAAAGGATTTACCGTCCATTCGGTATTACAAATACG 98  
 DB 160 ATTAGCTTAAATTCGTCGACTAAGAGGAAAAAACTTTAATTCGTTTTTACTAAAAA 101  
 QY 99 GCTGATTACGGAATTCGCCAAGCCGTAAAGATTTACGTTAATTCGCAACAGATAACC 158  
 DB 100 GAGGGTTGAAAATTTGGTATTAACGTAATTTGGTTAACTATATATTAACACACCAAAAT 41  
 QY 159 AAGAAA 165  
 DB 40 AAAAAATA 34

RESULT 11  
 US-09-557-884-1  
 ; Sequence 1, Application US/09557884  
 ; Patent No. 6506581  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Fleischmann et al.  
 ; TITLE OF INVENTION: The Nucleotide sequence of  
 ; the Haemophilus influenzae Rd Genome, Fragments  
 ; Thereof, and Uses Thereof  
 ; NUMBER OF SEQUENCES: 1  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Human Genome Sciences, Inc.  
 ; STREET: 9410 Key West Avenue  
 ; CITY: Rockville  
 ; STATE: MD  
 ; COUNTRY: USA  
 ; ZIP: 20850  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: 3 1/2 inch diskette  
 ; COMPUTER: Dell Pentium  
 ; OPERATING SYSTEM: MS DOS v6.22  
 ; SOFTWARE: ASCII Text  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/557,884  
 ; FILING DATE: 25-Apr-2000  
 ; CLASSIFICATION: <Unknown>  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/476,102  
 ; FILING DATE: JUN-5-1995  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Michelle S. Marks  
 ; REGISTRATION NUMBER: 41,971  
 ; REFERENCE/DOCKET NUMBER: PB186P3  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 301-309-8504  
 ; TELEFAX: 301-309-8439  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1830121 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
 US-09-557-884-1

Query Match 8.8%; Score 31; DB 4; Length 1830121;  
 Best Local Similarity 52.8%; Pred. No. 47;  
 Matches 67; Conservative 0; Mismatches 60; Indels 0; Gaps 0;  
 QY 40 ATTTCATAAAGTTATTGGAAGAAAAGGATTTACCGTCCATTCGGTATTTCACAAATACGG 99

Db 1105186 ATATTATTCAATTAATTAATTTGGTTATGGTTGACCTATATTAGATGCT 1105245  
QY 100 CTGATTACGGAATCCCAAGCCGTAAGAAATTTACGTTAATTCGAAACAGATAACCA 159  
Db 1105246 CTACTGTGTTGTTCCACAATCTAGAACAGATTTTCCTTAATTCGAAATTAATAGTG 1105305  
QY 160 AAGAAA 166  
Db 1105306 AACATAA 1105312  
RESULT 12  
US-09-643-990A-1  
; Sequence 1, Application US/09643990A  
; Patent No. 6528289  
; GENERAL INFORMATION:  
; APPLICANT: Robert D. Fleischmann  
; Mark D. Adams  
; Owen White  
; Hamilton O. Smith  
; J. Craig Venter  
; TITLE OF INVENTION: The Nucleotide sequence of  
; the Haemophilus influenzae Rd Genome, Fragments  
; Thereof, and Uses Thereof  
; NUMBER OF SEQUENCES: 1  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville,  
; STATE: MD  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3 1/2 inch diskette  
; COMPUTER: Dell Pentium  
; OPERATING SYSTEM: MS DOS v6.22  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/643,990A  
; FILING DATE: 23-Aug-2000  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/487,429  
; FILING DATE: 1995-06-07  
; APPLICATION NUMBER: 08/426,787  
; FILING DATE: 1995-04-21  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kenley K. Hoover  
; REGISTRATION NUMBER: 40,302  
; REFERENCE/DOCKET NUMBER: PB186P1C1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 301-610-5790  
; TELEFAX: 310-309-8439  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1830121 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-643-990A-1  
Query Match 8.8%; Score 31; DB 4; Length 1830121;  
Best Local Similarity 52.8%; Pred. No. 47;  
Matches 67; Conservative 0; Mismatches 60; Indels 0; Gaps 0;  
QY 40 ATTCATAGATTTTGGAGAAAGGATTTACGCCATTCGGTATTCACATACGG 99  
Db 1105186 ATATTATTCAATTAATTAATTTGGTTATGGTTGACCTATATTAGATGCT 1105245  
QY 100 CTGATTACGGAATCCCAAGCCGTAAGAAATTTACGTTAATTCGAAACAGATAACCA 159

Db 1105246 CTACTGTGTTGTTCCACAATCTAGAACAGATTTTCCTTAATTCGAAATTAATAGTG 1105305  
QY 160 AAGAAA 166  
Db 1105306 AACATAA 1105312  
RESULT 13  
US-09-734-674-3/c  
; Sequence 3, Application US/09734674  
; Patent No. 6498022  
; GENERAL INFORMATION:  
; APPLICANT: WEI, Ming-Hui et al  
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,  
; TITLE OF INVENTION: AND USES THEREOF  
; FILE REFERENCE: CL001018  
; CURRENT APPLICATION NUMBER: US/09/734,674  
; CURRENT FILING DATE: 2000-12-13  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 202001  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)...(202001)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-734-674-3  
Query Match 8.7%; Score 30.8; DB 4; Length 202001;  
Best Local Similarity 54.4%; Pred. No. 25;  
Matches 62; Conservative 0; Mismatches 52; Indels 0; Gaps 0;  
QY 2 GCGAATTCGAGCAGATGAAGAAAGCAGCTTGATAATTTCAATAAGTTATTGAAGA 61  
Db 3209 GGAAGTCTGACAATAAACAAATACATTAAATGAATATATTTTCATGAGTACACTAGA 3150  
QY 62 AAAGGATTTACCGTCCATTCGGTATTCACATAGCGCTGATCGGATTC 115  
Db 3149 AAAGCAAGCACCCCTACATTTCTAATATATCATCATATTCATGAGGTATACC 3096  
RESULT 14  
US-07-712-833A-1  
; Sequence 1, Application US/07712833A  
; Patent No. 5175101  
; GENERAL INFORMATION:  
; APPLICANT: GOTZ, Friedrich  
; APPLICANT: SEEBER, Stefan  
; TITLE OF INVENTION: RECOMBINANT RESTRICTION ENZYME SAU3AI  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Armstrong, Nikaido, Marmelstein, Kubovcik &  
; ADDRESSEE: Murray  
; STREET: 1725 K Street N.W., Suite 1000  
; CITY: Washington D.C.  
; COUNTRY: United States of America  
; ZIP: 20006  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/712,833A  
; FILING DATE: 19910610  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: DE P 4018441.2  
; FILING DATE: 08-JUN-1990  
; ATTORNEY/AGENT INFORMATION:



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 15, 2003, 00:09:34 ; Search time 199.897 Seconds  
(without alignments)  
5771.357 Million cell updates/sec

Title: US-09-928-457-83

Perfect score: 353

Sequence: 1 CGGGAATTCAGCAGCAATG.....ACGGAGGAACCGTTGGCTTT 353

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 2169961 seqs, 134102185 residues

Total number of hits satisfying chosen parameters: 4339922

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:\*
- 12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*
- 13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:\*
- 14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:\*
- 15: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*
- 16: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*
- 17: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	353	100.0	353	10	US-09-928-457-83
2	43.4	12.3	4519	9	US-09-956-004-6
3	37.2	10.5	557	12	US-10-027-632-201983
4	37.2	10.5	557	12	US-10-027-632-201984
5	37.2	10.5	557	12	US-10-027-632-201985
6	37.2	10.5	557	12	US-10-027-632-201986
7	37.2	10.5	557	13	US-10-027-632-201987
8	37.2	10.5	557	13	US-10-027-632-201988
9	37.2	10.5	557	13	US-10-027-632-201989
10	37.2	10.5	557	13	US-10-027-632-201990
11	35.6	10.1	13084	12	US-10-027-632-201986
12	35.4	10.0	935	10	US-09-879-536-626
13	33.6	9.5	17136	14	US-10-114-170-158
14	33	9.3	960	9	US-09-939-980-46
15	33	9.3	2401	13	US-10-036-492-14
16	33	9.3	2434	13	US-10-036-492-9

Sequence 27, Appl  
Sequence 505, App  
Sequence 224164,  
Sequence 224164,  
Sequence 90983, A  
Sequence 317596,  
Sequence 90983, A  
Sequence 317596,  
Sequence 436, App  
Sequence 7, Appl  
Sequence 7, Appl  
Sequence 87, Appl  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 102, App  
Sequence 102, App  
Sequence 34, Appl  
Sequence 201070,  
Sequence 220599,  
Sequence 220600,  
Sequence 220599,  
Sequence 201070,  
Sequence 220599,  
Sequence 220600,  
Sequence 193599,  
Sequence 193600,  
Sequence 193601,  
Sequence 193602,  
Sequence 193599,  
Sequence 193600,

2512 13 US-10-036-492-27  
628 11 US-09-871-161-505  
528 12 US-10-027-632-224164  
528 13 US-10-027-632-224164  
558 12 US-10-027-632-90983  
558 12 US-10-027-632-317596  
558 13 US-10-027-632-90983  
558 13 US-10-027-632-317596  
9155 12 US-10-311-455-436  
10809 12 US-09-960-858-7  
10809 12 US-09-960-870-7  
271990 14 US-10-195-144-87  
580073 12 US-10-205-220-1  
640681 10 US-09-790-988-1  
1584 9 US-09-925-299-102  
1584 11 US-09-925-299-102  
7434 10 US-09-070-927A-34  
598 12 US-10-027-632-201070  
598 12 US-10-027-632-220599  
598 12 US-10-027-632-220600  
598 13 US-10-027-632-201070  
598 13 US-10-027-632-220599  
657 12 US-10-027-632-193599  
657 12 US-10-027-632-193600  
657 12 US-10-027-632-193601  
657 12 US-10-027-632-193602  
657 13 US-10-027-632-193599  
657 13 US-10-027-632-193600

#### ALIGNMENTS

#### RESULT 1

US-09-928-457-83  
; Sequence 83, Application US/09928457  
; Patent No. US20020164603A1

GENERAL INFORMATION:  
APPLICANT:  
TITLE OF INVENTION: DNA, specific proteins and peptides  
TITLE OF INVENTION: of the Neisseria meningitidis species bacteria, method  
TITLE OF INVENTION: for obtaining them and their biological application.

NUMBER OF SEQUENCES: 99

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patentin Release #1.0, Version #1.30 (OEB)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/928,457

FILING DATE: 2001-08-14

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/214,759

FILING DATE: 199-12-10

INFORMATION FOR SEQ ID NO: 83:

SEQUENCE CHARACTERISTICS:

LENGTH: 353 base pairs

TYPE: nucleotide

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: NO

US-09-928-457-83

Query Match 100.0%; Score 353; DB 10; Length 353;

Best Local Similarity 100.0%; Pred No. 5e-93; 0; Indels 0; Gaps 0;

Matches 353; Conservative 0; Mismatches 0;

QY 1 CGGGAATTCAGCAGCAATGAAAGAAAGCAGCGTTGATAATTCATAAAGTTATTGGAAG 60

DB 1 CGGGAATTCAGCAGCAATGAAAGAAAGCAGCGTTGATAATTCATAAAGTTATTGGAAG 60

QY 61 AAAAAGGATTTACCGTCATTTTCGGTATTCACAAATACGGCTGATTACGGAAATTCGCCAA 120  
 DB 61 AAAAAGGATTTACCGTCATTTTCGGTATTCACAAATACGGCTGATTACGGAAATTCGCCAA 120  
 QY 121 GCCGTAAAGATTTACGTTTAAATTCGAAACAGAAATTAACCAAGAAAGCTGGAACCCAGTCA 180  
 DB 121 GCCGTAAAGATTTACGTTTAAATTCGAAACAGAAATTAACCAAGAAAGCTGGAACCCAGTCA 180  
 QY 181 AGTATTCGGGCAACCGCTTACGGTACCGATGTTTTGGGAATGGAATTCGTTTCCCAA 240  
 DB 181 AGTATTCGGGCAACCGCTTACGGTACCGATGTTTTGGGAATGGAATTCGTTTCCCAA 240  
 QY 241 CATTATTCGAGACACCAAGACGAAACGGATTTTATCATAGCTGTGCGGGAATTTATCTG 300  
 DB 241 CATTATTCGAGACACCAAGACGAAACGGATTTTATCATAGCTGTGCGGGAATTTATCTG 300  
 QY 301 ATATCACTTGAACGATTGGCTTGATCTAAATTAACGAGGAACCGTTGGCTTT 353  
 DB 301 ATATCACTTGAACGATTGGCTTGATCTAAATTAACGAGGAACCGTTGGCTTT 353

## RESULT 2

US-09-956-004-6/c  
 ; Sequence 6, Application US/09956004  
 ; Patent No. US20020072595A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Patrick J. Dillon et al.  
 ; TITLE OF INVENTION: Nucleotide Sequences of Escherichia coli Pathogenicity Islands  
 ; FILE REFERENCE: PB324D1  
 ; CURRENT APPLICATION NUMBER: US/09/956,004  
 ; CURRENT FILING DATE: 2001-09-20  
 ; PRIOR APPLICATION NUMBER: 08/976,259  
 ; PRIOR FILING DATE: 1997-11-21  
 ; PRIOR APPLICATION NUMBER: 60/061,953  
 ; PRIOR FILING DATE: 1997-10-14  
 ; PRIOR APPLICATION NUMBER: 60/031,626  
 ; PRIOR FILING DATE: 1996-11-22  
 ; NUMBER OF SEQ ID NOS: 142  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 6  
 ; LENGTH: 4519.

; TYPE: DNA  
 ; ORGANISM: Escherichia coli  
 ; FEATURE:  
 ; NAME/KEY: (3483)..(3483)  
 ; LOCATION: (3483)..(3483)  
 ; OTHER INFORMATION: n equals a, t, g, or c  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (3487)..(3487)  
 ; OTHER INFORMATION: n equals a, t, g, or c  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (4292)..(4292)  
 ; OTHER INFORMATION: n equals a, t, g, or c  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (4318)..(4318)  
 ; OTHER INFORMATION: n equals a, t, g, or c  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (4323)..(4323)  
 ; OTHER INFORMATION: n equals a, t, g, or c

## US-09-956-004-6

Query Match 12.3%; Score 43.4; DB 9; Length 4519;  
 Best Local Similarity 55.7%; Pred. No. 0.041; 66; Indels 0; Gaps 0;  
 Matches 83; Conservative 0; Mismatches

QY 21 AAAGAAGCAGCGCTTCATAATTTCAATTAAGTTATTGGAAGAAAAAGGATTTACCGTCCAT 80  
 DB 979 AAAGAAGCGCCATTTTCAGGAGTTTATTAGCTACTTAAAGAGTTAGAGTAACTATATA 920  
 QY 81 TTCGTTATTCAAATACGGCTGATTACGGAATTCCTCCGAAACCGGTAAAGATTTACGTTA 140  
 DB 919 TCTTTTATAGCCAATGCTGAGAACTATGGGATTCCTCCCAAGAAAGAAAGACTCGTGCTC 860

QY 141 ATTGCAAAAGATAAACCAGAAAGAAAGCT 169  
 DB 859 TTAGCTAGTCAGTAGGTAAAGTTACCT 831

## RESULT 3

US-10-027-632-201983/c  
 ; Sequence 201983, Application US/10027632  
 ; Publication No. US20030204075A9  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wang, David G.  
 ; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
 ; FILE REFERENCE: 108827.129  
 ; CURRENT APPLICATION NUMBER: US/10/027,632  
 ; CURRENT FILING DATE: 2002-04-30  
 ; PRIOR APPLICATION NUMBER: US 60/218,006  
 ; PRIOR FILING DATE: 2000-07-12  
 ; PRIOR APPLICATION NUMBER: US 60/198,676  
 ; PRIOR FILING DATE: 2000-04-20  
 ; PRIOR APPLICATION NUMBER: US 60/193,483  
 ; PRIOR FILING DATE: 2000-03-29  
 ; PRIOR APPLICATION NUMBER: US 60/185,218  
 ; PRIOR FILING DATE: 2000-02-24  
 ; PRIOR APPLICATION NUMBER: US 60/167,363  
 ; PRIOR FILING DATE: 1999-11-23  
 ; PRIOR APPLICATION NUMBER: US 60/156,358  
 ; PRIOR FILING DATE: 1999-09-28  
 ; PRIOR APPLICATION NUMBER: US 60/146,002  
 ; PRIOR FILING DATE: 1999-08-09  
 ; NUMBER OF SEQ ID NOS: 325720  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 201983  
 ; LENGTH: 557  
 ; TYPE: DNA  
 ; ORGANISM: Human  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (1)...(557)  
 ; OTHER INFORMATION: n = A,T,C or G

Query Match 10.5%; Score 37.2; DB 12; Length 557;  
 Best Local Similarity 51.2%; Pred. No. 1;  
 Matches 87; Conservative 0; Mismatches 83; Indels 0; Gaps 0;  
 QY 25 AAAGCAGGCTTGATTAATTTCAATAAGTTATTGGAAAGAAAAAGGATTTACCGTCCATTTCG 84  
 DB 406 AGAGAAATGCCAATTAATTCGAAACATTAGTTTAAAGACTAGGTGGATACCTAGCATTCAT 347  
 QY 85 GTATTCAATACGCTGATTACGGATTCCTCCCAAGCCGTAAGATTTACGTTAATTG 144  
 DB 346 GTAATGTTAATCTCTACTTTAGAAAATGTCAAAATTAATAAAGAACTATATTTTATT 287  
 QY 145 CAAACAGATAAACCAAGAAAAAGCTGGAAACCAAGTCAAGTATTCGGGCAAA 194  
 DB 286 TAAACATAAATAACATCATTCATGATACAGTAGACTAATAGGACTAA 237

## RESULT 4

US-10-027-632-201984/c  
 ; Sequence 201984, Application US/10027632  
 ; Publication No. US20030204075A9  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wang, David G.  
 ; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
 ; FILE REFERENCE: 108827.129  
 ; CURRENT APPLICATION NUMBER: US/10/027,632  
 ; CURRENT FILING DATE: 2002-04-30  
 ; PRIOR APPLICATION NUMBER: US 60/218,006  
 ; PRIOR FILING DATE: 2000-07-12



```
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 201984
; LENGTH: 557
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(557)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-201984

Query Match
Best Local Similarity 10.5%; Score 37.2; DB 12; Length 557;
Matches 87; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

QY 25 AAAGCAGGCTTCATATTTTCATAAAGTTTATTTGGAAGAAAAGGATTTACCGTCCATTTCG 84
Db 406 AGAGATGCCAATTAATTGCAACATTAGTTTAAAGACTAGGTGGATACCTAGCATTCAT 347
QY 85 GTATTCACAATACGGCTGATTACGGAATTCGCCAAGCCGTAAGATTTACGTTAATTG 144
Db 346 GTAATGTTAATCTCTRACTTTAGAAAATGTCAAAATTAATAAAGAACTATATTTATT 287
QY 145 CAACACAGATAACCAAGAAAAGCTGGAACAGTCAGTCAAGTATTCGGGCAA 194
Db 286 TAAACATAAAAATAACATCATTCATGAATACAGTACACTAATAGGACTAA 237
```

```
RESULT 5
US-10-027-632-201985/c
; Sequence 201985, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 10827,129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 201985
; LENGTH: 557
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(557)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-201985

Query Match
Best Local Similarity 10.5%; Score 37.2; DB 12; Length 557;
Matches 87; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

QY 25 AAAGCAGGCTTCATATTTTCATAAAGTTTATTTGGAAGAAAAGGATTTACCGTCCATTTCG 84
Db 406 AGAGATGCCAATTAATTGCAACATTAGTTTAAAGACTAGGTGGATACCTAGCATTCAT 347
QY 85 GTATTCACAATACGGCTGATTACGGAATTCGCCAAGCCGTAAGATTTACGTTAATTG 144
Db 346 GTAATGTTAATCTCTRACTTTAGAAAATGTCAAAATTAATAAAGAACTATATTTATT 287
QY 145 CAACACAGATAACCAAGAAAAGCTGGAACAGTCAGTCAAGTATTCGGGCAA 194
Db 286 TAAACATAAAAATAACATCATTCATGAATACAGTACACTAATAGGACTAA 237
```

```
; NAME/KEY: misc_feature
; LOCATION: (1)...(557)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-201985

Query Match
Best Local Similarity 10.5%; Score 37.2; DB 12; Length 557;
Matches 87; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

QY 25 AAAGCAGGCTTCATATTTTCATAAAGTTTATTTGGAAGAAAAGGATTTACCGTCCATTTCG 84
Db 406 AGAGATGCCAATTAATTGCAACATTAGTTTAAAGACTAGGTGGATACCTAGCATTCAT 347
QY 85 GTATTCACAATACGGCTGATTACGGAATTCGCCAAGCCGTAAGATTTACGTTAATTG 144
Db 346 GTAATGTTAATCTCTRACTTTAGAAAATGTCAAAATTAATAAAGAACTATATTTATT 287
QY 145 CAACACAGATAACCAAGAAAAGCTGGAACAGTCAGTCAAGTATTCGGGCAA 194
Db 286 TAAACATAAAAATAACATCATTCATGAATACAGTACACTAATAGGACTAA 237
```

```
RESULT 6
US-10-027-632-201986/c
; Sequence 201986, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 10827,129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 201986
; LENGTH: 557
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(557)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-201986
```

```
Query Match
Best Local Similarity 10.5%; Score 37.2; DB 12; Length 557;
Matches 87; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

QY 25 AAAGCAGGCTTCATATTTTCATAAAGTTTATTTGGAAGAAAAGGATTTACCGTCCATTTCG 84
Db 406 AGAGATGCCAATTAATTGCAACATTAGTTTAAAGACTAGGTGGATACCTAGCATTCAT 347
QY 85 GTATTCACAATACGGCTGATTACGGAATTCGCCAAGCCGTAAGATTTACGTTAATTG 144
Db 346 GTAATGTTAATCTCTRACTTTAGAAAATGTCAAAATTAATAAAGAACTATATTTATT 287
QY 145 CAACACAGATAACCAAGAAAAGCTGGAACAGTCAGTCAAGTATTCGGGCAA 194
```

```

Db      286 TAAACATAAAAATACATCATTCATGATACAGTAGACTAATAGGACTAA 237
US-10-027-632-201983/c
; Sequence 201983, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027.632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 201983
; LENGTH: 557
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(557)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-201983
Query Match      10.5%; Score 37.2; DB 13; Length 557;
Best Local Similarity 51.2%; Pred. No. 1;
Matches 87; Conservative 0; Mismatches 83; Indels 0; Gaps 0;
QY      25 AAAGCAGCGCTTGATTAATTCATTAAGTTATTGGAGAAAAAGGATTTACCGTCCATTTCG 84
Db      406 AGAGAATGCCAATTAATTCGAACATTAGTTTAAAGACTAGGTGGATACCTAGCATTCAT 347
QY      85 GTATTACAATACGGCTGATTACGGAATTCCTCCAAAGCCGTAAGATTTACGTTAATTG 144
Db      346 GTAATGTTAACTCTRACTTTAGAAATGTCAAATTTATAATAAAAACTATATTTTATT 287
QY      145 CAACAGAAATACCAAGAAAGCTGGAACCAAGTCAAGTATTCGGGCAA 194
Db      286 TAAACATAAAAATACATCATTCATGATACAGTAGACTAATAGGACTAA 237
RESULT 9
US-10-027-632-201985/c
; Sequence 201985, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027.632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 201985
; LENGTH: 557
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(557)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-201985
Query Match      10.5%; Score 37.2; DB 13; Length 557;
Best Local Similarity 51.2%; Pred. No. 1;
Matches 87; Conservative 0; Mismatches 83; Indels 0; Gaps 0;
QY      25 AAAGCAGCGCTTGATTAATTCATTAAGTTATTGGAGAAAAAGGATTTACCGTCCATTTCG 84
Db      406 AGAGAATGCCAATTAATTCGAACATTAGTTTAAAGACTAGGTGGATACCTAGCATTCAT 347
QY      85 GTATTACAATACGGCTGATTACGGAATTCCTCCAAAGCCGTAAGATTTACGTTAATTG 144
Db      346 GTAATGTTAACTCTRACTTTAGAAATGTCAAATTTATAATAAAAACTATATTTTATT 287
QY      145 CAACAGAAATACCAAGAAAGCTGGAACCAAGTCAAGTATTCGGGCAA 194
Db      286 TAAACATAAAAATACATCATTCATGATACAGTAGACTAATAGGACTAA 237
RESULT 8
US-10-027-632-201984/c
; Sequence 201984, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027.632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 201984
; LENGTH: 557
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(557)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-201984
Query Match      10.5%; Score 37.2; DB 13; Length 557;
Best Local Similarity 51.2%; Pred. No. 1;
Matches 87; Conservative 0; Mismatches 83; Indels 0; Gaps 0;
QY      25 AAAGCAGCGCTTGATTAATTCATTAAGTTATTGGAGAAAAAGGATTTACCGTCCATTTCG 84
Db      406 AGAGAATGCCAATTAATTCGAACATTAGTTTAAAGACTAGGTGGATACCTAGCATTCAT 347
QY      85 GTATTACAATACGGCTGATTACGGAATTCCTCCAAAGCCGTAAGATTTACGTTAATTG 144
Db      346 GTAATGTTAACTCTRACTTTAGAAATGTCAAATTTATAATAAAAACTATATTTTATT 287
QY      145 CAACAGAAATACCAAGAAAGCTGGAACCAAGTCAAGTATTCGGGCAA 194
Db      286 TAAACATAAAAATACATCATTCATGATACAGTAGACTAATAGGACTAA 237

```

```
Best Local Similarity 51.2%; Pred. No. 1;
Matches 87; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

QY 25 AAAGCAGGCTTGATTAATTCATAAAGTTATTGGAAGAAAAGGATTTACCGTCCATTTCG 84
Db 406 AGAAGATGCCAATTAATTCGAACATTTAGTTTAAAGACTAGTGATACCTAGCAATTCAT 347
QY 85 GTATTCAATACAGGCTGATTCAGGAATTCGCCAAAGCGGTAAAGATTTACGTTAAATG 144
Db 346 GTAATGTTAACTCTTACTTTAGAAAATGTCAAAATTAATAAATAACTTATATTTTATT 287
QY 145 CAAACAGAATAACCAAGAAAAGCTGGACCACTCAAGTATTTCGGGCAAA 194
Db 286 TAAACATAAAATAACATCATTCATGATACAGTAGACTAATAGACTAA 237

RESULT 10
US-10-027-632-201986/c
; Sequence 201986, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: Polymorphisms in the Human Genome
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 201986
; LENGTH: 557
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(557)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-201986

Query Match 10.5%; Score 37.2; DB 13; Length 557;
Best Local Similarity 51.2%; Pred. No. 1;
Matches 87; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

QY 25 AAAGCAGGCTTGATTAATTCATAAAGTTATTGGAAGAAAAGGATTTACCGTCCATTTCG 84
Db 406 AGAAGATGCCAATTAATTCGAACATTTAGTTTAAAGACTAGTGATACCTAGCAATTCAT 347
QY 85 GTATTCAATACAGGCTGATTCAGGAATTCGCCAAAGCGGTAAAGATTTACGTTAAATG 144
Db 346 GTAATGTTAACTCTTACTTTAGAAAATGTCAAAATTAATAAATAACTTATATTTTATT 287
QY 145 CAAACAGAATAACCAAGAAAAGCTGGACCACTCAAGTATTTCGGGCAAA 194
Db 286 TAAACATAAAATAACATCATTCATGATACAGTAGACTAATAGACTAA 237

RESULT 11
US-10-311-455-2152/c
; Sequence 2152, Application US/10311455
; Publication No. US20030143606A1
```

```
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Deter
; FILE REFERENCE: cytosine methylation
; CURRENT APPLICATION NUMBER: US/10/311,455
; PRIOR FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 2152
; LENGTH: 13084
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-2152

Query Match 10.1%; Score 35.6; DB 12; Length 13084;
Best Local Similarity 54.6%; Pred. No. 12;
Matches 71; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 37 ATAAATTCATAAAGTTATTGGAAGAAAAGGATTTACCGTCCATTTCGGTATTCACAATA 96
Db 9929 AAATATTCATAAATAATATACAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 9870
QY 97 CGGCTGATTTACGGAATTCGCCAAAGCGGTAAAGATTTACGTTAAATTCGAAACAGAAATA 156
Db 9869 TCAATATATAACAAACCCGCTAAATAACCAAAACATAAACTTTATCTCTCTAAATAA 9810
QY 157 CCAAGAGAAA 166
Db 9809 CCAAAAAAAA 9800

RESULT 12
US-09-879-536-626/c
; Sequence 626, Application US/09879536
; Patent No. US20020144298A1
; GENERAL INFORMATION:
; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carroll III, Eddie
; APPLICANT: Catino, Theodore J.
; APPLICANT: Derti, Adnan
; APPLICANT: Ford, Donna M.
; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.
; APPLICANT: Schlegel, Robert
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/879,536
; PRIOR FILING DATE: 2001-09-21
; PRIOR APPLICATION NUMBER: US 60/088,801
; PRIOR FILING DATE: 1998-06-10
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 626
; LENGTH: 935
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
```

Db 13604 TTCTTCCTGTTATTGCCAGCAAAAGGATATAACAATACAAAAGCAATTGATTTCTTCT 13663  
 QY 102 GATTCAGGAATCCCAAGCGTAAAGATTACGTTAATTGC 145  
 Db 13664 CATGATAAAATTTCCCAACGGCGTTAATGATGTTATTGAATCC 13707

RESULT 14  
 US-09-939-980-46/c  
 ; Sequence 46, Application US/09939980  
 ; Patent No. US20020082234A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Black, Michael  
 ; Burnham, Martin  
 ; Hodgson, John  
 ; Knawles, David  
 ; Lonetto, Michael  
 ; Pratt, Julie  
 ; Reichard, Richard  
 ; Rosenberg, Martin  
 ; Ward, Judith  
 ; TITLE OF INVENTION: No. US20020082234A1el Prokaryotic Polynucleotides,  
 ; NUMBER OF SEQUENCES: 534  
 ; CORRESPONDENCE ADDRESS:  
 ; STREET: 709 Swedeland Road  
 ; CITY: King of Prussia  
 ; STATE: PA  
 ; COUNTRY: USA  
 ; ZIP: 19406-0939  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; OPERATING SYSTEM: DOS  
 ; SOFTWARE: FastSeq for Windows Version 2.0  
 ; CURRENT APPLICATION DATA: US/09/939,980  
 ; FILING DATE: 27-Aug-2001  
 ; CLASSIFICATION: <Unknown>  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/936,165  
 ; FILING DATE: <Unknown>  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Gimmi, Edward R  
 ; REGISTRATION NUMBER: 38,891  
 ; REFERENCE/DOCKET NUMBER: P50549  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 610-270-4478  
 ; TELEFAX: 610-270-5090  
 ; TELEX: <Unknown>  
 ; INFORMATION FOR SEQ ID NO: 46:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 960 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: Genomic DNA  
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 46:  
 ; US-09-939-980-46

Query Match 9.3%; Score 33; DB 9; Length 960;  
 Best Local Similarity 51.7%; Pred. No. 22;  
 Matches 75; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 21 AAGAAGCAGCGCTTGATTAATTTTATGAGATTTTGCAGAGAAAAGGATTACCGTCCAT 80  
 Db 882 AAGAAGTAAATGATTTTTCAGCAAGATAAAATGAAATCATCGGAATTCCTA 823  
 QY 81 TTCGGTATTCACAAATACGCTGATTACGGAATTCCTCAAGAGATTTACGTTA 140  
 Db 822 TTGTGATGCCITGAATGCTAAATCGTAAATGTAATGTAATTAATTAATTT 763

LOCATION: (1)...(935)  
 OTHER INFORMATION: n = A,T,C or G  
 US-09-879-536-626  
 Query Match 10.0%; Score 35.4; DB 10; Length 935;  
 Best Local Similarity 60.0%; Pred. No. 4.4;  
 Matches 57; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 129 AGATTTCGTTAATTCAGCAATACCAAGAAAGCTGGAACAGTCAAGTATTTCG 188  
 Db 342 AAATTTTNCACCCCAAGGAAANCCAAAGCCTTAATTGGACCAATTGATTTACG 283  
 QY 189 GCGAAGCGCTTACGGTAGCGGATTTTGGGAAT 223  
 Db 282 GCCCAAGGCTTTGGTACCGCAAGCTTCGGGAT 248

RESULT 13  
 US-10-114-170-158  
 ; Sequence 158, Application US/10114170  
 ; Publication No. US20030023075A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Blattner, Frederick R.  
 ; Burland, Valerie  
 ; Perna, Nicole T.  
 ; Plunkett, Guy  
 ; Welch, Rod  
 ; TITLE OF INVENTION: No. US20030023075A1el Sequences of E. coli O157  
 ; NUMBER OF SEQUENCES: 265  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Quarles & Brady  
 ; STREET: 1 South Pinckney Street  
 ; CITY: Madison  
 ; STATE: WI  
 ; COUNTRY: US  
 ; ZIP: 53701-2113  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.44mb storage  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Word Perfect 8.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/10/114,170  
 ; FILING DATE: 01-Apr-2002  
 ; CLASSIFICATION: <Unknown>  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 09/453,702  
 ; FILING DATE: 03-DEC-1999  
 ; APPLICATION NUMBER: 60/110,955  
 ; FILING DATE: 04-DEC-1998  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Seay, Nicholas J.  
 ; REGISTRATION NUMBER: 27386  
 ; REFERENCE/DOCKET NUMBER: 960296.95017  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (608) 251-5000  
 ; TELEFAX: (608) 251-9166  
 ; INFORMATION FOR SEQ ID NO: 158:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 17136  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 158:  
 ; US-10-114-170-158

Query Match 9.5%; Score 33.6; DB 14; Length 17136;  
 Best Local Similarity 57.7%; Pred. No. 53;  
 Matches 60; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 42 TTCATAAGCTATTGCAAGAAAAGGATTTACCGTCCATTTGGTATTCAAAATACGGCT 101

QY 141 ATTGCAACAGATAACCAAGAA 165  
 Db 762 ATAGAAACATAAAATTCACAAA 738

RESULT 15:

US-10-036-492-14  
 ; Sequence 14, Application US/10036492  
 ; Publication No. US20020164757A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: HENERLY, ADRIANA  
 ; APPLICANT: FERREIRA, PAULO  
 ; APPLICANT: ROMBAUTS, STEPHANE  
 ; TITLE OF INVENTION: PLANT DNA REPLICATION MODULATING PROTEINS  
 ; FILE REFERENCE: 217943USOXCONT  
 ; CURRENT APPLICATION NUMBER: US/10/036,492  
 ; PRIOR FILING DATE: 2002-01-07  
 ; PRIOR APPLICATION NUMBER: EP 99202214.5  
 ; PRIOR FILING DATE: 1999-07-05  
 ; NUMBER OF SEQ ID NOS: 35  
 ; SOFTWARE: Patent in version 3.1  
 ; SEQ ID NO 14  
 ; LENGTH: 2401  
 ; TYPE: DNA  
 ; ORGANISM: Arabidopsis thaliana  
 US-10-036-492-14

Query Match 9.3%; Score 33; DB 13; Length 2401;  
 Best Local Similarity 69.2%; Pred. No. 33;  
 Matches 45; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 91 ACAATACGGCTGATTACGGAATCCCAAGCCGTAAGATTACGTTAATTCGCAACA 150  
 Db 532 ACCATAGACCAGATTACAGATTCTGATAGGCTTAAAGATACAGGTTATCGCAACA 591  
 QY 151 GAATA 155  
 Db 592 GAACA 596

Search completed: November 15, 2003, 08:32:04  
 Job time : 201.997 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: November 14, 2003, 23:56:24 ; Search time 1195.2 Seconds  
(without alignments)  
9241.609 Million cell updates/sec

Title: US-09-928-457-88  
Perfect score: 270  
Sequence: 1 AATTTCGACGGGAGG.....CTCGAACTCAATTCATT 270

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2889711 seqs, 2045481396 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_hg.\*
- 3: gb\_in.\*
- 4: gb\_ov.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vl.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_on.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vl.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_mus.\*
- 34: em\_htg\_pln.\*
- 35: em\_htg\_rdt.\*
- 36: em\_htg\_mam.\*
- 37: em\_htg\_vrt.\*
- 38: em\_sy.\*
- 39: em\_atgo\_hum.\*
- 40: em\_atgo\_mus.\*
- 41: em\_atgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	270	100.0	270	6	A68917	A68917 Sequence 88
2	270	100.0	270	6	BD063017	BD063017 DNA and s
3	264.2	97.9	271	1	AF169464	AF169464 Neisseria
C 4	254.8	94.4	349061	1	NMA222491	AL162753 Neisseria
C 5	253.2	93.8	858	6	AX024111	AX024111 Sequence
C 6	251.6	93.2	858	6	AX024058	AX024058 Sequence
C 7	242.2	89.7	270	1	AF169428	AF169428 Neisseria
C 8	133.4	49.4	852	1	AF169459	AF169459 Neisseria
C 9	121.8	45.1	269	1	AF169451	AF169451 Neisseria
C 10	121.8	45.1	269	6	A68906	A68906 Sequence 77
C 11	121.8	45.1	269	6	BD063006	BD063006 DNA and s
C 12	48	17.8	267334	1	AE016997	AE016997 Chlamydo
C 13	44.6	16.5	3930	1	LM0302030	AJ302030 Listeria
C 14	43.4	16.1	92413	1	AL646075	AL646075 Ralstonia
C 15	43	15.9	4884	1	HGD1IM	X55141 H.giganteus
C 16	42.6	15.8	2578	3	AY051762	AY051762 Drosophil
C 17	42.6	15.8	3170	3	AB049434	AB049434 Drosophil
C 18	42.6	15.8	20175	2	AC019990	AC019990 Drosophil
C 19	42.6	15.8	170498	3	AC008094	AC008094 Drosophil
C 20	42.6	15.8	244757	3	AE003678	AE003678 Drosophil
C 21	41.4	15.3	123098	9	AC008799	AC008799 Homo sapi
C 22	41.4	15.3	151031	9	AC008923	AC008923 Homo sapi
C 23	39.6	14.7	139746	2	AC079599	AC079599 Homo sapi
C 24	39.6	14.7	158531	9	AC025575	AC025575 Homo sapi
C 25	39.6	14.7	192773	2	AC092917	AC092917 Homo sapi
C 26	39.6	14.7	193010	2	AC041044	AC041044 Homo sapi
C 27	39.2	14.5	164845	2	AC121528	AC121528 Mus muscu
C 28	38.4	14.2	209037	2	AL845360	AL845360 Mus muscu
C 29	38.4	14.2	146087	2	AC141709	AC141709 Apis mell
C 30	38.4	14.2	195985	2	EX119979	EX119979 Danio rer
C 31	38.4	14.2	213419	2	EX511083	EX511083 Danio rer
C 32	38.2	14.1	3276	5	BC044098	BC044098 Xenopus l
C 33	38.2	14.1	6068	6	AX344796	AX344796 Sequence
C 34	37.6	13.9	15455	1	AE002263	AE002263 Chlamydo
C 35	37.6	13.9	16903	1	AE001662	AE001662 Chlamydia
C 36	37.6	13.9	110000	6	AR310754	Continuation (10 o
C 37	37.6	13.9	181168	2	EX293540	EX293540 Danio rer
C 38	37.6	13.9	212999	2	AC111972	AC111972 Rattus no
C 39	37.6	13.9	274806	2	AC126582	AC126582 Rattus no
C 40	37.6	13.9	32585	1	AP02548	AP02548 Chlamydo
C 41	37.4	13.9	192199	2	AC139312	AC139312 Bos tauru
C 42	37.4	13.9	239901	2	AC125552	AC125552 Rattus no
C 43	37.4	13.9	262424	2	AC103010	AC103010 Rattus no
C 44	37.2	13.8	112063	9	AC091485	AC091485 Homo sapi
C 45	37.2	13.8	180488	2	AC125116	AC125116 Mus muscu

ALIGNMENTS

RESULT 1  
A68917  
LOCUS A68917  
DEFINITION Sequence 88 from Patent WO9802547.  
ACCESSION A68917  
VERSION A68917.1 GI:4759836  
KEYWORDS  
SOURCE unidentified  
ORGANISM unclassified  
REFERENCE 1 (bases 1 to 270)  
AUTHORS Nassif, X., Tinsley, C., Achtman, M., Ruelle, J., Vinals, C. and Merker, P.  
TITLE DNA AND SPECIFIC PROTEINS OR PEPTIDES OF THE NEISSERIA MENINGITIDIS SPECIES BACTERIA, METHOD FOR OBTAINING THEM AND THEIR BIOLOGICAL

linear PAT 06-MAY-1999

APPLICATIONS  
Patent: WO 9802547-A 88 22-JAN-1998;  
INSR NAT SANTE RECH MED (FR)  
Other publication FR 2751000 19980116.  
Location/Qualifiers  
1..270  
/organism="unidentified"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32644"  
BASE COUNT 72 a 67 c 42 g 89 t  
ORIGIN

Query Match 100.0%; Score 270; DB 6; Length 270;  
Best Local Similarity 100.0%; Pred. No. 1.1e-58;  
Matches 270; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AATCTTCGCGACGGGAGGCTTGTTTCTCTCCCTTCTGTTCCGACCGATTCTCAATA 60  
DB 1 AATCTTCGCGACGGGAGGCTTGTTTCTCTCCCTTCTGTTCCGACCGATTCTCAATA 60

QY 61 AAAATCATTGATTTTCATCGAAGTTTCATTCCTATACCAATTCCTTAATAACGATTTTATG 120  
DB 61 AAAATCATTGATTTTCATCGAAGTTTCATTCCTATACCAATTCCTTAATAACGATTTTATG 120

QY 121 CTCGGTTTATCGAATAACCTTAACCTTCACCTTCGCTAGCACATCGTAGGCAATTCGC 180  
DB 121 CTCGGTTTATCGAATAACCTTAACCTTCACCTTCGCTAGCACATCGTAGGCAATTCGC 180

QY 181 TATCAACTCGGCAATCGCAGGAACAGTGTGCAATACAAATCTTTACACCCAAATGTTTGA 240  
DB 181 TATCAACTCGGCAATCGCAGGAACAGTGTGCAATACAAATCTTTACACCCAAATGTTTGA 240

QY 241 TTACGGTTGGCTCGAAACTCAATTTCAAAT 270  
DB 241 TTACGGTTGGCTCGAAACTCAATTTCAAAT 270

RESULT 3  
AF169464 271 bp DNA linear BCT 09-AUG-2000  
LOCUS Neisseria meningitidis strain Z2491 clone Em024 unknown sequence.  
DEFINITION AF169464  
ACCESSION AF169464  
VERSION AF169464.1 GI:9754672  
KEYWORDS  
SOURCE Neisseria meningitidis  
ORGANISM Neisseria meningitidis  
Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
Neisseriaceae; Neisseria.  
REFERENCE 1 (bases 1 to 271)  
AUTHORS Perrin,A., Nassif,X. and Tinsley,C.R.  
TITLE Identification of regions of the chromosome of Neisseria meningitidis and Neisseria gonorrhoeae which are specific to pathogenic Neisseriae  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 271)  
AUTHORS Perrin,A., Nassif,X. and Tinsley,C.R.  
TITLE Direct Submission  
JOURNAL Submitted (16-JUL-1999) Necker Medicine Faculty, INSERM U411, 156 rue de Vaugirard, Paris 75015, France  
FEATURES  
source  
1..271  
/organism="Neisseria meningitidis"  
/mol\_type="genomic DNA"  
/strain="Z2491"  
/db\_xref="taxon:487"  
/clone="Em024"  
BASE COUNT 72 a 67 c 42 g 89 t 1 others  
ORIGIN

Query.Match 97.9%; Score 264.2; DB 1; Length 271;  
Best Local Similarity 98.5%; Pred. No. 3.4e-57;  
Matches 266; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AATCTTCGCGACGGGAGGCTTGTTTCTCTCCCTTCTGTTCCGACCGATTCTCAATA 60  
DB 1 AATCTTCGCGACGGGAGGCTTGTTTCTCTCCCTTCTGTTCCGACCGATTCTCAATA 60

QY 61 AAAATCATTGATTTTCATCGAAGTTTCATTCCTATACCAATTCCTTAATAACGATTTTATG 120  
DB 61 AAAATCATTGATTTTCATCGAAGTTTCATTCCTATACCAATTCCTTAATAACGATTTTATG 120

APPLICATIONS  
Patent: WO 9802547-A 88 22-JAN-1998;  
INSR NAT SANTE RECH MED (FR)  
Other publication FR 2751000 19980116.  
Location/Qualifiers  
1..270  
/organism="unidentified"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32644"  
BASE COUNT 72 a 67 c 42 g 89 t  
ORIGIN

Query Match 100.0%; Score 270; DB 6; Length 270;  
Best Local Similarity 100.0%; Pred. No. 1.1e-58;  
Matches 270; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AATCTTCGCGACGGGAGGCTTGTTTCTCTCCCTTCTGTTCCGACCGATTCTCAATA 60  
DB 1 AATCTTCGCGACGGGAGGCTTGTTTCTCTCCCTTCTGTTCCGACCGATTCTCAATA 60

QY 61 AAAATCATTGATTTTCATCGAAGTTTCATTCCTATACCAATTCCTTAATAACGATTTTATG 120  
DB 61 AAAATCATTGATTTTCATCGAAGTTTCATTCCTATACCAATTCCTTAATAACGATTTTATG 120

QY 121 CTCGGTTTATCGAATAACCTTAACCTTCACCTTCGCTAGCACATCGTAGGCAATTCGC 180  
DB 121 CTCGGTTTATCGAATAACCTTAACCTTCACCTTCGCTAGCACATCGTAGGCAATTCGC 180

QY 181 TATCAACTCGGCAATCGCAGGAACAGTGTGCAATACAAATCTTTACACCCAAATGTTTGA 240  
DB 181 TATCAACTCGGCAATCGCAGGAACAGTGTGCAATACAAATCTTTACACCCAAATGTTTGA 240

QY 241 TTACGGTTGGCTCGAAACTCAATTTCAAAT 270  
DB 241 TTACGGTTGGCTCGAAACTCAATTTCAAAT 270

RESULT 2  
BD063017 270 bp DNA linear PAT 27-AUG-2002  
LOCUS DNA and specific proteins or peptides of the Neisseria meningitidis species bacteria, method for obtaining them and their biological applications.  
DEFINITION BD063017.1 GI:22608620  
ACCESSION JP 2001504684-A/79.  
VERSION JP 2001504684-A/79.  
KEYWORDS unidentified  
SOURCE unidentified  
ORGANISM unclassified.  
REFERENCE 1 (bases 1 to 270)  
AUTHORS Nassif,X., Tinsley,C., Achtman,M., Ruelle,J.L., Vinals,C. and Merker,P.  
TITLE DNA and specific proteins or peptides of the Neisseria meningitidis species bacteria, method for obtaining them and their biological applications  
JOURNAL Patent: JP 2001504684-A 79 10-APR-2001;  
INSTITUT NATIONAL DE LA SANTE ET DE LA RECHERCHE MEDICALE, MAX PLANCK GESELLSCHAFT ZUR FORDERUNG DER WISSENSCHAFTEN EV BERLIN, SMITHKLINE BEECHAM  
PN JP 2001504684-A/79  
COMMENT  
PD 10-APR-2001  
PF 11-JUL-1997 JP 1998505685  
PR 12-JUL-1996 FR 96/08768  
PI XAVIER NASSIF, COLIN TINSLEY, MARK ACHTMAN, JEAN LOUIS RUELLE, PI CARLA VINALS,  
PI PETRA MERKER  
PC C12N15/31, C07K14/22, C07K16/12, A61K9/095, C12Q1/68, G01N33/53 CC  
Strandedness: Single;  
CC Topology: Linear;  
FH Key Location/Qualifiers.  
1..270  
/organism="unidentified"



QY 121 CTCGGGTTTATCGAATAAAGCTTAACCTTCCACTTCCTCGTAGCACATGCATCGTAGGCATTGCG 180  
 Db |||||  
 121 CTCGGGTTTATCGAATAAAGCTTAACCTTCCACTTCCTCGTAGCACATGCATCGTAGGCATTGCG 180  
 QY 181 TATCAACTCGGCAATCGCAGGACAGCTGTGCGAATACAACTTTACACCCAAATGTTGGA 240  
 Db |||||  
 181 TATCAACTCGGCAATCGCAGGACAGCTGTGCGAATACAACTTTACACCCAAATGTTGGA 240  
 QY 241 TTACGGTTGCTCGAAACTCAATTTCAAT 270  
 Db |||||  
 241 TTACGGTTGCTCGAAACTCAATTTCAAT 270

## RESULT 4

NMA222491/c NMA222491 349061 bp DNA linear BCT 02-SEP-2002  
 LOCUS Neisseria meningitidis serogroup A strain Z2491 complete genome;  
 DEFINITION segment 2/7.  
 ACCESSION AL162753 AL157959  
 VERSION AL162753.2 GI:7379120  
 KEYWORDS

## SOURCE

ORGANISM Neisseria meningitidis Z2491  
 Neisseria meningitidis Z2491  
 Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
 Neisseriaceae; Neisseria.

## REFERENCE

AUTHORS Parkhill, J., Achtman, M., James, K. D., Bentley, S. D., Churcher, C.,  
 Klee, S. R., Morelli, G., Basham, D., Brown, D., Chillingworth, I.,  
 Davies, R. M., Davis, P., Devlin, K., Feltwell, T., Hamlin, N.,  
 Holtroyd, S., Jagers, K., Leath, S., Moule, S., Mungall, K.,  
 Quail, M. A., Rajandream, M. A., Rutherford, K. M., Simmonds, M.,  
 Skelton, J., Whitehead, S., Spratt, B. G. and Barrall, B. G.  
 Complete DNA sequence of a serogroup A strain of Neisseria  
 meningitidis Z2491.

## TITLE

Nature 404 (6777), 502-506 (2000)

## JOURNAL

MEDLINE 20222556

## PUBMED

10761919

## REFERENCE

2 (bases 1 to 349061)

## AUTHORS

Parkhill, J.

## TITLE

Submitted (30-MAR-2000) Submitted on behalf of the Neisseria  
 sequencing team, Sanger Centre, Wellcome Trust Genome Campus,  
 Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk

## JOURNAL

Notes:

Details of N. meningitidis sequencing at the Sanger Centre are  
 available on the World Wide Web.

(URL, [http://www.sanger.ac.uk/Projects/N\\_meningitidis/](http://www.sanger.ac.uk/Projects/N_meningitidis/)).

## FEATURES

## source

1..349061  
 /organism="Neisseria meningitidis Z2491"  
 /mol\_type="genomic DNA"  
 /strain="Z2491"  
 /db\_xref="taxon:122587"  
 /note="serogroup: A"  
 complement(24..206)  
 /note="AIR repeat; hmms hit to HMM ATR (1 - 183), score:  
 310.39"

## repeat\_unit

/label=ATR

## RBS

209..212

## gene

220..681

## CDS

/gene="NMA0368"

/note="NMA0368, probable integral membrane protein, len:  
 153 aa; contains four probable transmembrane domains"

/codon\_start=1

/transl\_table=1

/product="putative integral membrane protein"

/protein\_id="CAB83669.1"

/db\_xref="GI:7379121"

/translation="MQEQRKSSPPIVMLLVVALMTIASLNVAFYLGNGSMEGLTV

LIIGSIASIDIRYCAVANYVMLAAIIVLLALRKKVVPVHAAFWGLALVAFSVKAVYV  
 DEAGTSDIVRYGAGFYLWAAFAVAGISGTAGNKKRKAASAADTKNDV"  
 671..1492  
 /gene="hemK"  
 671..1492  
 /gene="hemK"  
 /note="NMA0369, hemK, HemK protein, len: 273 aa; similar  
 to e.g. HEMK\_ECOLI P3786 HEMK protein (277 aa), fasta  
 scores; E(): 0, 42.3% identity in 279 aa overlap. Contains  
 PS00092 N-6 Adenine-specific DNA methylases signature"  
 /codon\_start=1  
 /transl\_table=11  
 /product="HemK protein"  
 /protein\_id="CAB83670.1"  
 /db\_xref="GI:7379122"  
 /db\_xref="SPTREMBL:Q9JWH6"  
 /translation="MFIDKVLGSLKPKNEARMLOVSVYTRVQLITRGCEMPDPEV  
 RQADRRAQRLNAGEVAVYILGAREFVGRFTVPSVLIIPRTEHLVEAVLRLPEN  
 GRYVDLGTGSAVAVTVALLERPDFAVRASDISPALETAKNAADILGARVEAYGSHF  
 DTMPSEGGKWDIIIVSNPPYIENGDKHLSQSDLRFBQCIALTDFSDGLSCIRTIYAGQAP  
 DRLAEGGFLLEHGFQDGAARVGLAENGFGFVETILPDLAQLDRVILGKYNKHLK"  
 831..840  
 /gene="hemK"  
 /note="Core DNA uptake sequence: gccgtctgaa"  
 /label=DUS  
 1162..1171  
 /gene="hemK"  
 /note="Core DNA uptake sequence: gccgtctgaa"  
 /label=DUS  
 1187..1207  
 /gene="hemK"  
 /note="PS00092 N-6 Adenine-specific DNA methylases  
 signature"  
 /note="Core DNA uptake sequence: gccgtctgaa"  
 /label=DUS  
 1565..1568  
 1578..2966  
 /gene="NMA0370"  
 1578..2966  
 /gene="NMA0370"  
 /note="NMA0370, probable integral membrane protein, len:  
 462 aa; similar to hypothetical proteins e.g. Y325\_HAEIN  
 P44640 hypothetical protein HI0325 (450 aa), fasta\_scores;  
 E(): 0, 49.8% identity in 464 aa overlap"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative integral membrane protein"  
 /protein\_id="CAB83671.1"  
 /db\_xref="GI:7379123"  
 /db\_xref="SPTREMBL:Q9JWH5"  
 /translation="MNAVVVAVIVMLVLSLRVHVLSLTIGAFVGVAGVMPLOXIA  
 DAAGVSOAGIIPVENKGLGGAKIALSYAMLGAPAMATHSGLPQOLAGAVYKLNLR  
 GMPDSVRSRGEVVKWLLLSILLVMGMSQNIPIHIAIPIMVPLLVFNKIKDIR  
 RLIACVITFGLVTTNMFPGFAIFLNEILIGNTHSAAPQLDVKNINVMAMAIPAL  
 GMLAGLLAFVHKRPLRYOSNNADTGNADANRPPQSVYSLAAVAIVCAIQL  
 MYEDSLGMLGFAVFMGLVGNIRNDKANDVFEQGIKMMAVGFMIAACQGAAMVNA  
 TGHIPQLVESNVAIFGNSKMAALMLVVGLVLTWIGGSSFTLPIIAIYVPLCVGL  
 GSPSLATVAIVGTAGALGDAGSPASDSTLGPMTGLNADQGDHDIRDSVPIRTHNIP  
 LLIAGWIAAMVL"  
 2967..3191  
 /gene="slyX"  
 2967..3191  
 /gene="slyX"  
 /note="NMA0371, slyX, SLX protein homolog, len: 74 aa;  
 similar to SLX\_HAEIN P44759 SLX protein homolog (73 aa),  
 fasta\_scores; E(): 0.0017, 33.8% identity in 74 aa  
 overlap, and SLX\_ECOLI F30857 SLX protein (72 aa), fasta  
 scores; E(): 0.24, 32.4% identity in 68 aa overlap"  
 /codon\_start=1  
 /transl\_table=11  
 /product="slyX protein homolog"  
 /protein\_id="CAB83672.1"

```

/db_xref="GI:7379124"
/db_xref="SPTREMBL:Q9JWH4"
/translation="MDAVQFPEHRIITEIQSLQEDVIAGLNMAVELRQLTDLQQA
QLRLYQKMDRNPDAQEPYSLRDEIPHY"
complement(3271..3423)
/gene="NMA0372"
complement(3271..3423)
/gene="NMA0372"
/notes="NMA0372, unknown, questionable CDS, len: 50 aa"
/codon_start=1
/transl_table=11
/product="very hypothetical protein NMA0372"
/protein_id="CAB83673.1"
/db_xref="GI:7379125"
/db_xref="SPTREMBL:Q9JWH3"
/translation="MFCRLSPGSDGFLLAVYEPVGFICLQGTAYFRSGGRKPHRI
KGIIISD"
misc_feature
3562..3571
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(3586..4356)
/gene="thif"
complement(3586..4356)
/gene="thif"
/notes="NMA0373, thif, probable Thif protein, len: 256 aa;
similar to e.g. THIF_ECOLI P30138 THIF protein (251 aa),
fasta scores; E(): 0, 43.1% identity in 246 aa overlap,
and MOEB_ECOLI P12282 molybdopterin biosynthesis MOEB
protein, (249 aa), fasta scores; E(): 0, 43.9% identity in
244 aa overlap (note that N.m. does not have orthologs of
any other molybdopterin biosynthesis proteins). Contains
Pfam match to entry PF00899 Thif_family, Thif family"
/codon_start=1
/transl_table=11
/product="Thif protein"
/protein_id="CAB83674.1"
/db_xref="GI:7379126"
/db_xref="SPTREMBL:Q9JWH2"
/translation="MTTEHNDDAFLRYSRHILLDEIGEOOKLSAAHILVGGG
GLGAALPLAASGILTIADSDIVELHNLQROVAEDGVGKLTKEALADLRHIN
HTVDRTINEKDCRLTGLVQADIVLDCNDYATQAVNQCQVQPLVSGAAR
FEGQAVYRPLDPSYACILFDGGSADGICSLFVGFPLVIGISTQAAEALKILL
DAGPQHGLRAVYALEGMQYFDLPNPECPVCGAER"
misc_feature
3787..3796
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(3847..4257)
/gene="thif"
/notes="Pfam match to entry PF00899 Thif_family, Thif
family, score 186.60, E-value 4.1e-52"
complement(4418..4427)
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
4471..4724
/gene="ppc"
4471..4724
/gene="ppc"
/EC_numbers="4.1.1.31"
/notes="NMA0374, ppc, phosphoenolpyruvate carboxylase, len:
917 aa; similar to many e.g. CAP2_RHOA Q32483
phosphoenolpyruvate carboxylase (EC 4.1.1.31) (936 aa),
fasta scores; E(): 0, 43.3% identity in 928 aa overlap.
Contains 2x Pfam match to entry PF00311 PEPcase,
Phosphoenolpyruvate carboxylase, PS00017 ATP/Grp-binding
site motif A (p-loop); and PS00393 Phosphoenolpyruvate
carboxylase active site 2"
/codon_start=1
/transl_table=11
/product="phosphoenolpyruvate carboxylase"
/protein_id="CAB83675.1"
Query Match 94.4%; Score 254.8; DB 1; Length 349061;
Best Local Similarity 98.9%; Pred. No. 9.1e-55;

```

```

Matches 267; Conservative 0; Mismatches 2; Indels 1; Gaps 1;
QY 1 AATCTTCGCGAGGGAGGCTGTTTTCCTCTCTGTCGACCGATTCTCAATA 60
DB 59503 AATCTTCGCGAGGGAGGCTGTTTTCCTCTCTGTCGACCGATTCTCAATA 59444
QY 61 AAAATCATTGATTCATCGAAGTTCATTCCTATACCATTTCTTAATACGATTTTATG 120
DB 59443 AAAATCATTGATTCATCGAAGTTCATTCCTATACCATTTCTTAATACGATTTTATG 59384
QY 121 CTCGGTTTATCGAATAACCTAACCTTCCTAGTACATGTCATCGTAGGATTTCGC 180
DB 59383 CTCGGTTTATCGAATAACCTAACCTTCCTAGTACATGTCATCGTAGGATTTCGC 59324
QY 181 TATCAACTCGGCAATCGGAGGAACAGTGTGGATACAACTTTACACCCAAATGTTTGA 240
DB 59323 TATCAACTCGGCAATCGGAGGAACAGTGTGGATAC-ATCTTTACACCCAAATGTTTGA 59265
QY 241 TTACGGTTGGCTCGAAACTCAATTTCAATT 270
DB 59264 TTACGGTTGGCTCGAAACTCAATTTCAATT 59235
RESULT 5
AX024111/c
LOCUS AX024111 858 bp DNA linear PAT 15-SEP-2000
DEFINITION Sequence 54 from Patent FR2785293.
ACCESSION AX024111
VERSION AX024111.1 GI:10184422
KEYWORDS
SOURCE Neisseria gonorrhoeae
ORGANISM Neisseria gonorrhoeae
Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
Neisseriaceae; Neisseria.
REFERENCE
1 Nassif, X., Tinsley, C., Aujame, L., Perrin, A., Rokbi, B.,
Bouchardon, A. and Renauld, M.G.
Patent: FR 2785293-A 54 05-MAY-2000;
PASTEUR MERIEUX SERUMS VACC (FR)
FEATURES
Location/Qualifiers
source 1..858
/organism="Neisseria gonorrhoeae"
/mol_type="genomic DNA"
/db_xref="taxon:485"
1..858
/notes="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAC08931.1"
/db_xref="GI:10184423"
/translation="MSEELKMSPEFTVIEHLGVKMSHTVPAIAELIANAYDACATE
VEVRLFDKPEHKIVIKDNGIMSFDEINDFVLRGNRRREKQASPCGRITGRKGLG
KLALFRLGNKIEISTIQGNRVFTFLDYAEIKKSERIYQPEFQKESVPKNTENGTTIT
LTFLPKQGVPLDNDVYVHLSRLFPADPKIKVSLNGSEPRIIDGNLKNLVTQFFE
WEYQDLATNISLSKSPQYEGYSGLIQGFITTEKPLNKNMGIPLFANGRMVNPPEF
FIDSESSH"
BASE COUNT 302 a 151 c 178 g 227 t
ORIGIN
Query Match 93.8%; Score 253.2; DB 6; Length 858;
Best Local Similarity 98.5%; Pred. No. 2.2e-54;
Matches 266; Conservative 0; Mismatches 3; Indels 1; Gaps 1;
QY 1 AATCTTCGCGAGGGAGGCTGTTTTCCTCTCTGTCGACCGATTCTCAATA 60
DB 282 AATCTTCGCGAGGGAGGCTGTTTTCCTCTCTGTCGACCGATTCTCAATA 223
QY 61 AAAATCATTGATTCATCGAAGTTCATTCCTATACCATTTCTTAATACGATTTTATG 120
DB 222 AAAATCATTGATTCATCGAAGTTCATTCCTATACCATTTCTTAATACGATTTTATG 163
QY 121 CTCGGTTTATCGAATAACCTAACCTTCCTAGTACATGTCATCGTAGGATTTCGC 180

```

Db 162 CTCGGTTTATCGAATAACCTAACCTCCACCTTCGAGACATGTCATCGTAGGCATTGCG 103  
 Qy 181 TATCAACTCGGAATCGAGGACAGTGTGGAAATACATCTTTACACCCAAATGTTTGA 240  
 Db 102 TATCAACTCGGAATCGAGGACAGTGTGGAAATAC-ATCTTTACACCCAAATGTTTGA 44  
 Qy 241 TTACGGTTGGCTCGAAACTCAATTTCAATT 270  
 Db 43 TTACGGTTGGCTCGAAACTCAATTTCAATT 14

RESULT 6  
 AX024058/c  
 LOCUS  
 DEFINITION  
 Sequence 1 from Patent FR2785293.  
 ACCESSION  
 AX024058  
 VERSION  
 AX024058.1 GI:10184370  
 KEYWORDS  
 ORGANISM  
 Neisseria meningitidis  
 Neisseria meningitidis  
 Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
 Neisseriaceae; Neisseria.

REFERENCE  
 1 Nassif, X., Tinsley, C., Aujame, L., Perrin, A., Rokbi, B.,  
 Bouchardon, A. and Renaud, M.G.  
 Patent: FR 2785293-A 1 05-MAY-2000;  
 PASTEUR MERIEUX SERUMS VACC (FR)

FEATURES  
 Location/Qualifiers  
 1..858  
 /organism="Neisseria meningitidis"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:487"  
 1..858  
 /note="unnamed protein product"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAC08905.1"  
 /db\_xref="GI:10184371"  
 /translation="MSSEKLMSPEPTVIEHLGVKMSHTVPAIAELIANAYDACATE  
 VEVLFPDPEHKIVKINGIGMSDFEINDFVLRICNRNBEKOASPCGRPTGKGLG  
 KALFLGLCNKEIETIGONERVFTLDVAEIRRSKGIYQEFERKESVESNIESGTT  
 LLETKKQYPLDNYVHLSPLFPDQDEPKIYSLNGSEPKIIDGNLKLDTVPOPE  
 WEYQDLATNIISSLSKSEQVEYSLGQKITEPKLNNKGIITFLLPGRWYNNMPEF  
 FIDSESHF"

BASE COUNT 297 a 145 c 182 g 234 t

CDS  
 Query Match 93.2%; Score 251.6; DB 6; Length 858;  
 Best Local Similarity 98.1%; Pred. No. 5.7e-54;  
 Matches 265; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

Qy 1 AATCTTCGCGACGGGAGGCTTTTCTTCTCCCTCTCTGTCGACCGATTCTCAATA 60  
 Db 282 AATCTTCGCGACGGGAGGCTTTTCTTCTCCCTCTCTGTCGACCGATTCTCAATA 223

Qy 61 AARATCATGTTTATCATCGAGTTTCATCTTATACCATATCTTTAATACGATTTATG 120  
 Db 222 AARATCATGTTTATCATCGAGTTTCATCTTATACCATATCTTTGATACGATTTATG 163

Qy 121 CTCGGTTTATCGAATAACCTAACCTTCACCTTCGTCAGCAGCATCGTCGAGGATTTCGC 180  
 Db 162 CTCGGTTTATCGAATAACCTAACCTTCACCTTCGTCAGCAGCATCGTCGAGGATTTCGC 103

Qy 181 TATCAACTCGGCATCGCAGGACAGTGTGGAAATACAAATCTTTACACCCAAATGTTTGA 240  
 Db 102 TATCAACTCGGCATCGCAGGACAGTGTGGAAATAC-ATCTTTACACCCAAATGTTTGA 44

Qy 241 TTACGGTTGGCTCGAAACTCAATTTCAATT 270  
 Db 43 TTACGGTTGGCTCGAAACTCAATTTCAATT 14

RESULT 7

AF169428  
 LOCUS  
 DEFINITION  
 Neisseria gonorrhoeae strain FA1090 clone Eg010 unknown sequence.  
 ACCESSION  
 AF169428  
 VERSION  
 AF169428.1 GI:9754636  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Neisseria gonorrhoeae  
 Neisseria gonorrhoeae  
 Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
 Neisseriaceae; Neisseria.

REFERENCE  
 1 (bases 1 to 270)  
 Perrin, A., Nassif, X. and Tinsley, C.R.  
 Identification of regions of the chromosome of Neisseria  
 meningitidis and Neisseria gonorrhoeae which are specific to  
 pathogenic Neisseriae  
 Unpublished

REFERENCE  
 2 (bases 1 to 270)  
 Perrin, A., Nassif, X. and Tinsley, C.R.  
 Direct Submission  
 Submitted (16-JUL-1999) Necker Medicine Faculty, INSERM U411, 156  
 rue de Vaugirard, Paris 75015, France  
 Location/Qualifiers  
 1..270  
 /organism="Neisseria gonorrhoeae"  
 /mol\_type="genomic DNA"  
 /strain="FA1090"  
 /db\_xref="taxon:485"  
 /clone="Eg010"

BASE COUNT 69 a 69 c 43 g 89 t

ORIGIN  
 Query Match 89.7%; Score 242.2; DB 1; Length 270;  
 Best Local Similarity 98.2%; Pred. No. 1.4e-51;  
 Matches 266; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

Qy 1 AATCTTCGCGACGGGAGGCTTTTCTTCTCCCTCTCTGTCGACCGATTCTCAAT 59  
 Db 1 AATCTTCGCGACGGGAGGCTTTTCTTCTCCCTCTCTGTCGACCGATTCTCAAT 60

Qy 60 AAAAAATCATGTTTATCATCGAGTTCATCTTATACCATATCTTTAATACGATTTAT 119  
 Db 61 AAAAAATCATGTTTATCATCGAGTTCATCTTATGTCATTTCTTTAATACGATTTAT 120

Qy 120 GTCTCCGGTTTATCGAATAACCTAACCTTCACCTTCGTCAGCAGCATCGTAGGCATTGCG 179  
 Db 121 GTCTCCGGTTTATCGAATAACCTAACCTTCACCTTCGTCAGCAGCATCGTAGGCATTGCG 180

Qy 180 CTATCAACTCGGCATCGCAGGACAGTGTGGAAATACATCTTTACACCCAAATGTTTGC 239  
 Db 181 CTATCAACTCGGCATCGCAGGACAGTGTGGAAATAC-ATCTTTACACCCAAATGTTTGC 239

Qy 240 ATTACGGTTGGCTCGAAACTCAATTTCAATT 270  
 Db 240 ATTACGGTTGGCTCGAAACTCAATTTCAATT 270

RESULT 8  
 AF169459/c  
 LOCUS  
 DEFINITION  
 Neisseria meningitidis strain Z2491 clone Cml30 unknown sequence.  
 ACCESSION  
 AF169459  
 VERSION  
 AF169459.1 GI:9754667  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Neisseria meningitidis  
 Neisseria meningitidis  
 Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
 Neisseriaceae; Neisseria.

REFERENCE  
 1 (bases 1 to 852)  
 Perrin, A., Nassif, X. and Tinsley, C.R.  
 Identification of regions of the chromosome of Neisseria  
 meningitidis and Neisseria gonorrhoeae which are specific to  
 pathogenic Neisseriae  
 Unpublished

```

REFERENCE 2 (bases 1 to 852)
AUTHORS Perrin,A., Nassif,X. and Tinsley,C.R.
TITLE Direct Submission
JOURNAL Submitted (16-JUL-1999) Necker Medicine Faculty, INSERM U411, 156
rue de Vaugirard, Paris 75015, France
FEATURES
    source
        1..852
            /organism="Neisseria meningitidis"
            /mol_type="genomic DNA"
            /strain="22491"
            /db_xref="taxon:487"
            /clone="Cm130"
BASE COUNT 272 a 154 c 206 g 220 t
ORIGIN
Query Match 49.4%; Score 133.4; DB 1; Length 852;
Best Local Similarity 98.6%; Pred. No. 9.2e-24;
Matches 145; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
QY 124 CGGTTTATCGAATAACCTTAACCTTCGTTCCGTTAGCAGATCGATCGTAGGCAATTCGCTAT 183
DB 852 CGGTTTATCGAATAACCTTAACCTTCGTTCCGTTAGCAGATCGATCGTAGGCAATTCGCTAT 793
QY 184 CAACCTGGCAATCGCAGAAACAGTGTGGCAATACAACTTTACACCCAAATGTCGATTA 243
DB 792 CAACCTGGCAATCGCAGAAACAGTGTGGCAATAC-ATCTTTACACCCAAATGTCGATTA 734
QY 244 CGGTTGGTCCGAAACTCAATTTCAATT 270
DB 733 CGGTTGGTCCGAAACTCAATTTCAATT 707

RESULT 9
LOCUS AF169451/c 269 bp DNA linear BCT 09-AUG-2000
DEFINITION Neisseria meningitidis strain Z2491 clone Cm040 unknown sequence.
ACCESSION AF169451
VERSION AF169451.1 GI:9754659
KEYWORDS
SOURCE
ORGANISM
    Neisseria meningitidis
    Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
    Neisseriaceae; Neisseria.
REFERENCE 1 (bases 1 to 269)
AUTHORS Perrin,A., Nassif,X. and Tinsley,C.R.
TITLE Identification of regions of the chromosome of Neisseria meningitidis and Neisseria gonorrhoeae which are specific to pathogenic Neisseriae
JOURNAL Unpublished
AUTHORS Perrin,A., Nassif,X. and Tinsley,C.R.
TITLE Direct Submission
JOURNAL Submitted (16-JUL-1999) Necker Medicine Faculty, INSERM U411, 156
rue de Vaugirard, Paris 75015, France
FEATURES
    source
        1..269
            /organism="Neisseria meningitidis"
            /mol_type="genomic DNA"
            /strain="Z2491"
            /db_xref="taxon:487"
            /clone="Cm040"
BASE COUNT 95 a 41 c 65 g 68 t
ORIGIN
Query Match 45.1%; Score 121.8; DB 1; Length 269;
Best Local Similarity 98.4%; Pred. No. 8.4e-21;
Matches 123; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 AATTTCTCGCAGCGGAGGCTTGTTTCTCCCTCTGTTCCGACCGATTCTCAATA 60
DB 125 AATTTCTCGCAGCGGAGGCTTGTTTCTCCCTCTGTTCCGACCGATTCTCAATA 66
QY 61 AAATCAATGATTTCATCGAAGTTCATTCCTATACCATTAATCTTTAATACGATTATG 120
DB 65 AAATCAATGATTTCATCGAAGTTCATTCCTATACCATTAATCTTTAATACGATTATG 6
QY 121 CTCCG 125
DB 5 CTCCG 1

RESULT 11
LOCUS BD063006/c 269 bp DNA linear PAT 27-AUG-2002
DEFINITION DNA and specific proteins or peptides of the Neisseria meningitidis species bacteria, method for obtaining them and their biological applications.
ACCESSION BD063006
VERSION BD063006.1 GI:22608609
KEYWORDS JF 2001504684-A/89.
SOURCE
ORGANISM
    Neisseria meningitidis
    Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
    Neisseriaceae; Neisseria.
REFERENCE 1 (bases 1 to 269)
AUTHORS Nassif,X., Tinsley,C., Achtman,M., Ruelle,J.L., Vinals,C. and
Merker,P.
TITLE DNA and specific proteins or peptides of the Neisseria meningitidis species bacteria, method for obtaining them and their biological applications
JOURNAL Patent: JP 2001504684-A 68 10-APR-2001;
INSTITUT NATIONAL DE LA SANTE ET DE LA RECHERCHE MEDICALE, MAX
PLANCK GESELLSCHAFT ZUR FORDERUNG DER WISSENSCHAFTEN EV BERLIN,

```

```

Db 65 AAATCAATGATTTCATCGAAGTTCATTCCTATACCATTAATCTTTAATACGATTATG 6
QY 121 CTCCG 125
DB 5 CTCCG 1

RESULT 10
LOCUS A68906/c 269 bp DNA linear PAT 06-MAY-1999
DEFINITION Sequence 77 from Patent WO9802547.
ACCESSION A68906
VERSION A68906.1 GI:4759825
KEYWORDS
SOURCE
ORGANISM
    Neisseria meningitidis
    Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
    Neisseriaceae; Neisseria.
REFERENCE 1 (bases 1 to 269)
AUTHORS Nassif,X., Tinsley,C., Achtman,M., Ruelle,J., Vinals,C. and
Merker,P.
TITLE DNA and specific proteins or peptides of the NEISSERIA MENINGITIDIS SPECIES BACTERIA, METHOD FOR OBTAINING THEM AND THEIR BIOLOGICAL APPLICATIONS
JOURNAL Patent: WO 9802547-A 77 22-JAN-1998;
INST NAT SANTE RECH MED (FR)
COMMENT Other publication FR 2751000 19980116.
FEATURES
    source
        1..269
            /organism="Neisseria meningitidis"
            /mol_type="genomic DNA"
            /db_xref="taxon:32644"
BASE COUNT 95 a 41 c 65 g 68 t
ORIGIN
Query Match 45.1%; Score 121.8; DB 6; Length 269;
Best Local Similarity 98.4%; Pred. No. 8.4e-21;
Matches 123; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 AATTTCTCGCAGCGGAGGCTTGTTTCTCCCTCTGTTCCGACCGATTCTCAATA 60
DB 125 AATTTCTCGCAGCGGAGGCTTGTTTCTCCCTCTGTTCCGACCGATTCTCAATA 66
QY 61 AAATCAATGATTTCATCGAAGTTCATTCCTATACCATTAATCTTTAATACGATTATG 120
DB 65 AAATCAATGATTTCATCGAAGTTCATTCCTATACCATTAATCTTTAATACGATTATG 6
QY 121 CTCCG 125
DB 5 CTCCG 1

RESULT 11
LOCUS BD063006/c 269 bp DNA linear PAT 27-AUG-2002
DEFINITION DNA and specific proteins or peptides of the Neisseria meningitidis species bacteria, method for obtaining them and their biological applications.
ACCESSION BD063006
VERSION BD063006.1 GI:22608609
KEYWORDS JF 2001504684-A/89.
SOURCE
ORGANISM
    Neisseria meningitidis
    Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
    Neisseriaceae; Neisseria.
REFERENCE 1 (bases 1 to 269)
AUTHORS Nassif,X., Tinsley,C., Achtman,M., Ruelle,J.L., Vinals,C. and
Merker,P.
TITLE DNA and specific proteins or peptides of the Neisseria meningitidis species bacteria, method for obtaining them and their biological applications
JOURNAL Patent: JP 2001504684-A 68 10-APR-2001;
INSTITUT NATIONAL DE LA SANTE ET DE LA RECHERCHE MEDICALE, MAX
PLANCK GESELLSCHAFT ZUR FORDERUNG DER WISSENSCHAFTEN EV BERLIN,

```

COMMENT	SMITHKLINE BEECHAM		gene	complement (251..421) /locus tag="CCA00790" complement (251..421) /locus tag="CCA00790" /note="identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AAP05531.1" /db_xref="GI:29834897" /translation="MEKGPICALFVLGEFHHSKNTVECKNMFESRKTFLYTLCKDFF IIVLCFRPSLFL" 451..1737 /locus tag="CCA00791" 451..1737 /locus tag="CCA00791" /note="similar to GB:X69464, PID:398210, and PID:398211; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="poly(A) polymerase family protein" /protein_id="AAP05532.1" /db_xref="GI:29834898" /translation="MVCDNKTILFRRGLELFRKISKSAPTIIVSAADHNILKIDPSPH ALSVKTLKAGHKAVIVGCIKIDLLNTPKDPDIDISTSAKPEIKAVKNCILVGR FRAHIRFENQIIEVSTFRSGADEDYLIKONLWGTAEEDVLRDFTINGFYDAE ETIIDYGVSDIQNRYLRTIGDFVFRKQDPVMRLKLLIARASFTVDPKLEALQ ECYELVKGAPAFEEILKMLGSGASSEFFKLVKYQMLEILFPYMDKAFRLKTLQ EQFSLCDVLDKNSVSKYNDYRHLQMAAFLLPFINVFNVRKHKRQPSLSLAFIDYI KNFLQFADSDKSSCKNFILIALVLQMYRLTPLVPTKKIHFFNRKFLNHRFSEA LSLEIIRSLVYPKLDKILAAVVRHYQALCKKELPS" 1751..3631 /gene="lpxB" /locus tag="CCA00792" 1751..3631 /gene="lpxB" /locus tag="CCA00792" /note="similar to GB:X69464, PID:398210, and PID:398211; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="lipid-A-disaccharide synthase" /protein_id="AAP05533.1" /db_xref="GI:29834899" /translation="MLPLYLVHVLPIGLIANLFTGSARTIOWLSEKBTAVVPKAF WLSSTGAVMIAGHGIQSPFMALLHGANLIVYFNALNIASSYKLSLITTLVILVLT LLVITDFFALAAIYIPYMEWASPNNFFHLPLEPPIYWHIVGCLGFTFSRFFIOWC YLEMNNHSTLPALFWOAGFVGGLAFIYFIRTPDPVNIILSYGCGLLPSLANRLIYKK SLPKFHSFSCFLSAGEPSGDTLGSDDLNIKELNPNHFCFVGGLFMRKEGLEPIIR MBEFQVSGLEVFCAVSLYKRYKLYKAILKENPETVFCIDFPDFHFLIKRLKCG YRGIYHVPQSIWAWRPNRKLILEKHLDTLLILPFEKEIFKDSPLKTIYLGHPLVK TIANFODCNAMKQOLEISQPSVALFGSREGDIFRNLOVQARAFSSSLAKSHOLV SSGNPKYDKLLELDKXECNNKIVPSKRYQLMRDCCALACKGTIVLEAALNTP TIVICLRFDIFLAKIIFKIFIPAYSLNITISVIFPEFVGGRHDFSEVAAAIID LIVATIGKSKQYACQQLKTMTENVITPEKCLQAIYAQNRFYLNKDFIKFHPKSS RA"
	PN JP 2001504684-A/68		CDS	
	PD 10-APR-2001			
	PF 11-JUL-1997 JP 1998505685			
	PR 12-JUL-1996 FR 96/08768			
	PI XAVIER NASSIF, COLIN TINSLEY, MARK ACHTMAN, JEAN LOUIS RUELE, PI CARLA VINALS,			
	PI PETRA MERKER			
	PC C12N15/31, C07K14/22, C07K16/12, A61K39/095, C12Q1/69, G01N33/53 CC			
	Strandedness: Single;			
	CC Topology: Linear;			
FEATURES	FH Key Location/Qualifiers.		gene	
	1..269		CDS	
	/organism="unidentified"			
	/mol_type="genomic DNA"			
	/db_xref="taxon:32644"			
	95 a 41 c 65 g 68 t			
	BASE COUNT			
	ORIGIN			
	Query Match 45.1%; Score 121.8; DB 6; Length 269;			
	Best Local Similarity 98.4%; Pred. No. 8.4e-21;			
RESULTS	Matches 123; Conservative 0; Mismatches 2; Indels 0; Gaps 0;			
	QY 1 AATCTTCGACGGGGAGGCTGTGTTTCTCCCTTCCTGTTCCGACCGATTCTCAATA 60			
	Db 125 AATCTTCGACGGGGAGGCTGTGTTTCTCCCTTCCTGTTCCGACCGATTCTCAATA 66			
	QY 61 AAAATCATGATTTCATGAAGTTCATTCCTATACATTATCTTTAATAACGATTTTATG 120			
	Db 65 AAAATCATGATTTCATGAAGTTCCTTATACATTATCTTTAATAACGATTTTATG 6			
	QY 121 CTCGG 125			
	Db 5 CTCGG 1			
	RESULT 12			
	AE016997			
	LOCUS			
REFERENCES	AE016997 267334 bp DNA linear BCT 14-APR-2003			
	Chlamydomophila caviae GPIC section 4 of 4 of the complete genome.			
	ACCESSION			
	AE016997.1 GI:29834896			
	VERSION			
	KEYWORDS			
	SOURCE			
	Chlamydomophila caviae GPIC			
	Chlamydomophila caviae GPIC			
	Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydomophila.			
AUTHORS	Read, I.D., Myers, G.S., Brunham, R.C., Nelson, W.C., Paulsen, I.T.,			
	Heidelberg, J., Holtzapfel, E., Khouri, H., Federova, N.B., Carty, H.A.,			
	Umayam, L.A., Haft, D.H., Peterson, J., Beanan, M.J., White, O.,			
	Salzberg, S.L., Hsia, R.C., McClarty, G., Rank, R.G., Bavoil, P.M. and			
	Fraser, C.M.			
	Genome sequence of Chlamydomophila caviae (Chlamydia psittaci GPIC):			
	examining the role of niche-specific genes in the evolution of the			
	Chlamydiaceae			
	Nucleic Acids Res. 31 (8), 2134-2147 (2003)			
	22569155			
JOURNAL	MEDLINE			
	PUBMED			
	12682364			
	REFERENCE			
	2 (bases 1 to 267334)			
	Read, T., Myers, G., Brunham, R., Nelson, W., Paulsen, I.,			
	Heidelberg, J., Holtzapfel, E., Khouri, H., Federova, N., Carty, H.,			
	Umayam, L., Haft, D., Peterson, J., Beanan, M., White, O., Salzberg, S.,			
	Hsia, R.-C., McClarty, G., Rank, R., Bavoil, P. and Fraser, C.			
	Direct Submission			
TITLE	Submitted (29-OCT-2002) The Institute for Genomic Research, 9712			
	Medical Center Dr. Rockville, MD 20850, USA			
	Location/Qualifiers			
	1..267334			
	/organism="Chlamydomophila caviae GPIC"			
	/mol_type="genomic DNA"			
	/isolate="GPIC"			
	/db_xref="taxon:227941"			
	FEATURES			
	source			
FEATURES	complement (3942..4700)		gene	complement (3942..4700) /locus tag="CCA00793" complement (3942..4700) /locus tag="CCA00793" /note="similar to GP:6382503; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AAP05534.1" /db_xref="GI:29834900" /translation="MRVHAMNVSLSVNPESTTNKFMHRFKKCLLSHTAAVAVGIV LILASVALIAYASLLICSLIFTSALIGVALISLGVKYVRYTLKIDPVDYSHERA GYGAMIKSLLEIQIDQDRVNEQAGDIEKQLEQVYKTLHSHITALGEQFKWKNQO KEWVCAHFSRAYERELKAVKLYEANKIEQORLOIPEELKVARERAVQILEWVQL RLSRDLOVAAANMIOFLQEEAEKVPKPRSRSASI" complement (4728..5849)
	/locus tag="CCA00793"		CDS	
	complement (3942..4700)			
	/locus tag="CCA00793"			
	/note="similar to GP:6382503; identified by sequence similarity; putative"			
	/codon_start=1			
	/transl_table=11			
	/product="hypothetical protein"			
	/protein_id="AAP05534.1"			
	/db_xref="GI:29834900"			









```

/function="miscellaneous; hypothetical/global homology"
/notes="Product confidence : hypothetical
Gene name confidence : hypothetical
Predicted by Codon usage
Predicted by Homology
Predicted by Framed"
/codon_start=1
/evidence=not experimental
/transl_table=11

Query Match      16.1%; Score 43.4; DB 1; Length 92413;
Best Local Similarity 49.8%; Pred. No. 0.96;
Matches 110; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

QY 6 TTCCGACGCGGAGGCTGTTTCTCTCCCTCTGTCGACCGCATTCCTCAATAAATAAT 65
Db 82192 TGCGAGACGAGTACCTTCTTCGCTCTCTTTCAGCTCGGCATATACAGACTAGAT 82133

QY 66 CATGTATTCATCGAAGTTCATCCATACCATATTCATTAACGATTTATGCTCCG 125
Db 82132 CAGCGAAGACCTCGCGCATTCATCCATTAACCGTCTCTCGAATTTATTTCTTCAAAC 82073

QY 126 GTTATTCGAATAACCACTTCACCTTCGTCAGCAGATCATCGTAGGATTCGGTATCA 185
Db 82072 GTGGGCAATCCGCTCTGAAATACTACAGAGTTGCGTCAGCGTCATAGGATAGATCA 82013

QY 186 ACTCGGCAATCGCAGGAAAGAGTGTGCGAATACAACTTTTAC 226
Db 82012 ATTCCGCAAGCGAAGCAGGTTGCGGATAGATTCATCC 81972

RESULT 15
HGDIIIM HGDIIIM 4884 bp DNA linear BCT 30-JUN-1993
DEFINITION H.giganteus DIIM gene for methyltransferase.
ACCESSION X55141 S65046
VERSION X55141.1 GI:48771
KEYWORDS methyltransferase.
SOURCE Herpetosiphon aurantiacus
ORGANISM Herpetosiphon aurantiacus
Bacteria; Chloroflexi; Herpetosiphonales; Herpetosiphonaceae;
Herpetosiphon.
Croeger M.
Direct Submission
Submitted (27-NOV-1990) M. Kroeger, INST MIKRO- UND
MOLEKULARBIOLOGIE, FRANKFURTER STRASSE 107, D-W-6300 GIESSEN, FRG
2 (bases 1 to 4884)
Dusterhoft A. and Kroeger M.
Cloning, sequence and characterization of
m5C-methyltransferase-encoding gene, hgdIIM (GTCGAC), from
Herpetosiphon giganteus strain Hpa2
Gene 106 (1), 87-92 (1991)
JOURNAL MEDLINE 92039068
PUBMED 1937045
COMMENT See also X55137-X55143.
FEATURES
    source      Location/Qualifiers
                1..4884
                /organism="Herpetosiphon aurantiacus"
                /mol_type="genomic DNA"
                /strain="Hpa2"
                /db_xref="taxon:65"
                884..1300
                /gene="orf 15"
                884..1300
                /gene="orf 15"
                /codon_start=1
                /transl_table=11
                /protein_id="CAA38940.1"
                /db_xref="GI:48772"
                /db_xref="SWISS-PROT:P25279"
                /translation="MIATGAYLWTIREAIGLCRDVNAHEAGTNNVQIMRIKGBSIDIR
                GSLLSVRAVNFNAEHIAQLFIMLVATEEDGRNLAIWINRETTSAIDFIAVYKXD

```

```

NKVSALKLIQOLEALDPSSLDRLLLGYQSLLDNR"
1749..2813
/gene="hgdIIM"
CDS
1749..2813
/gene="hgdIIM"
/codon_start=1
/transl_table=11
/product="methyltransferase"
/protein_id="CAA38941.1"
/db_xref="GI:48773"
/db_xref="SWISS-PROT:P25265"
/translation="MVGAVIDLCFVGGLTHGLILEGVLGADINDPSCKYAEQNN
RTRFEKISGEVDRELNALYFNQHKILVGCAPQDFSQYTKSRGTQWLLTEFS
RUIRIEPIIISSENVPEVTFENRGEVFNFIQLEQLGYHSHVSHVCPDYGIPQOR
DLVLFAAKQGVKIIPTHTPENYTRVDVIGSLATNYSGHWEGSMHAASLEDI
NURRIQHSVPGGTWADWPELLIAECHKESGESYGVYGRMWDKVAPTITTCNGYV
NGRFQHPQDRAISLREAAALQTPRSYQFAPEGQLKTKTVSRIGNAVPALGRVIA
KSIKFLGLEHERQVRVRII"
2782..4617
/gene="orf 68"
CDS
2782..4617
/gene="orf 68"
/codon_start=1
/transl_table=11
/protein_id="CAA38942.1"
/db_xref="GI:48774"
/db_xref="SWISS-PROT:P25280"
/translation="MSDSGYELSPSLNHLSEGLYSNILRFYKLVANAWDADATE
VSINRQDEIVIQNGIGSIEDANTKFLRIGHQKREDSANTI SGRHVMGRKGIGILA
IFGIANIAEVYCKDGVPHGLIHKGDIERGISDVTLYRPSVPEALSTESGTKII
LREIKSSICNAEKLRTLARRFTIINNKNFVSIINNTINDNDROYLVKQVFLWL
GGESKYANFTKLLKKSPEITNLGDSGMTVKWIGTVRPSDIPNEQPVHIFYAHG
XMQEDILLIDITDAGLYROYILGRIEADFMDSDDDEDIITSRQRIKQTDPRYLKLE
YKQIDIMVIASITWTKLRKEIPSNPKKEEVNDSLSKDSSEKENVINASSDSSTATE
NASSDSSTANASSETNDGEVDSNFFDDIPSPPPKQEIATFAFMKNLVKNSN
IPNQMKNILYDIQAAVAYKGFKACIVMLGALLEGVMTGITQRTDVLVLYALPQ
VFKPLSDLGPRNPKFADRTVLAQYIGTTFSPQCKEIELCVQGTNKLGVDILOTVRN
SIHPGSVLKMKQLARFNHQAGVYIAKLHEIINLVILWNPSP"
BASE COUNT 1500 a 882 c 1055 g 1447 t
ORIGIN

Query Match      15.9%; Score 43; DB 1; Length 4884;
Best Local Similarity 56.8%; Pred. No. 1.2;
Matches 79; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 29 TCTTCCTCTCTCTCGACCGATTCTCAAATAAAATCATTCATTTCATCGAAGTTCATT 88
Db 3026 TCTTCCTCTCTTGATGCGCCATTCTGAGATTTTAGTATTAGCATCTCTCATTCACATT 2967

QY 89 CCTATACCATTTATCTTTAATACGATTTTATGCTCCGGTTTATCGAATAACCTTACTCC 148
Db 2966 CCTATACCATTTATCTTGATGACTATTTTCATCTTCGTATATTATATAGATACCTCAGTT 2907

QY 149 ACTTCGCTAGCACATGCAT 167
Db 2906 GCATCAGCATCCCAAGCAT 2888

Search completed: November 15, 2003, 03:34:04
Job time : 1199.2 secs

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 14, 2003, 23:55:19 ; Search time 174.258 Seconds  
(without alignments)  
4182.570 Million cell updates/sec

Title: US-09-928-457-88

Perfect score: 270

Sequence: 1 ATTCTCCGACGGGGGG.....CTCGAATCAATTCATT 270

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_19Jun03.\*

- 1: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*
- 2: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*
- 3: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*
- 4: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*
- 5: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*
- 6: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*
- 7: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*
- 8: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*
- 9: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*
- 10: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*
- 11: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*
- 12: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*
- 13: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*
- 14: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*
- 15: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*
- 16: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*
- 17: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*
- 18: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*
- 19: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*
- 20: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*
- 21: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*
- 22: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*
- 23: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*
- 24: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*
- 25: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	270	100.0	270	19	AAV03596
2	270	100.0	270	21	AAAL5393
3	253.2	93.8	855	25	AB237755
4	253.2	93.8	858	21	AAAL5322
5	253.2	93.8	891	25	AB241353
6	251.6	93.2	858	21	AAAL5296
7	121.8	45.1	269	19	AAV03585
8	121.8	45.1	269	21	AAAL5389
					Neisseria meningit
					Genomic fragment o
					N gonorrhoeae nuc
					DNA encoding a pol
					N. gonorrhoeae nuc
					DNA encoding a pol
					Neisseria meningit
					Genomic fragment o

C 9	42.6	15.8	4951	23	ABL13122	Drosophila melanog
C 10	42.6	15.8	5395	23	ABL15458	Drosophila melanog
C 11	38.2	14.1	6068	24	ABN80204	Human chemically m
C 12	37.6	13.9	1230025	20	AAAG1990	Nucleotide sequenc
C 13	37	13.7	7160	20	AAAG6191	E. coli J96 pathog
C 14	36	13.3	4519	19	AAV31192	DNA encoding Esche
C 15	36	13.3	11165	21	AAAL5186	Drosophila melanog
C 16	35.4	13.1	7529	23	ABL04486	Lettuce resistance
C 17	34.6	12.8	1437	19	AAV44212	Lettuce pest resis
C 18	34.6	12.8	1437	24	ABX67809	Lettuce pest resis
C 19	34.6	12.8	5829	24	ABX67855	Lettuce pest resis
C 20	34.6	12.8	5933	24	ABK39977	Human chemically p
C 21	34.6	12.8	6013	24	AA561264	Human gene regulat
C 22	34.6	12.8	6013	24	ABK31360	Signal transductio
C 23	34.6	12.8	7980	19	AAV44243	Lettuce resistance
C 24	34.6	12.8	10151	24	ABL34433	Human immune syste
C 25	34.4	12.7	543	22	ABA61500	Human foetal liver
C 26	34.4	12.7	543	22	ABA29225	Probe #7691 for ge
C 27	34.4	12.7	543	22	AAK09799	Human brain expres
C 28	34.4	12.7	543	22	AAK35693	Human bone marrow
C 29	34.4	12.7	543	22	AA141408	Probe #10094 used
C 30	34.4	12.7	543	23	ABK35412	Human liver single
C 31	34.4	12.7	543	24	ABK09972	Human genome-deriv
C 32	34.2	12.7	1881	24	ABQ70528	Listeria monocytog
C 33	34.2	12.7	2365589	24	ABA90521	Genomic sequence o
C 34	34.2	12.7	2944528	24	ABA03041	Listeria monocytog
C 35	34	12.6	15767	24	ABL33206	Human immune syste
C 36	34	12.6	15767	24	ABL34552	Human metastasis a
C 37	33.8	12.5	1327	24	ABL99913	Human secretory po
C 38	33.4	12.4	285	21	AAK95132	Cat flea head and
C 39	33.4	12.4	285	21	AAK95358	Cat flea head and
C 40	33.2	12.3	13807	18	AAAT6328	Exopolysaccharide
C 41	33.2	12.3	14602	18	AAAT73236	Streptococcus ther
C 42	33.2	12.3	111309	20	AAAX20250	Borrelia burgdorfe
C 43	33.2	12.3	910715	20	AAAX20248	Borrelia burgdorfe
C 44	33	12.2	5360	24	ABL92297	Chemically created
C 45	33	12.2	5360	24	ABL49366	Human polynucleoti

#### ALIGNMENTS

RESULT 1  
AAV03596  
ID AAV03596 standard; DNA; 270 BP.  
XX  
AC AAV03596;  
XX  
DT 22-OCT-1998 (first entry)  
XX  
DE Neisseria meningitidis DNA sequence E24.

XX N. gonorrhoeae; N. lactamica; chromosome 22491; region 1; region 2;  
XX region 3; pathogenicity; blood-brain barrier; diagnosis; infection;  
XX meningitis; ss.

XX Neisseria meningitidis.

XX Neisseria meningitidis.

XX WO9802547-A2.

PD 22-JAN-1998.

XX 11-JUL-1997; 97WO-FR01295.

XX 12-JUL-1996; 96FR-0008768.

XX (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.

XX (BLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.

XX (SMIK ) SMITHKLINE BEECHAM.

XX Nassif X, Tinsley C, Achtman M, Merker P, Ruelle J;

XX Vinals C;

DR WPI; 1998-110594/10.

XX Genes present in *Neisseria meningitidis* but not other *Neisseria*

PT species - and related host cells, RNA, anti-sense sequences,

PT polypeptide(s) and antibodies, useful for diagnosing *Neisseria*

PT meningitidis infection and in protective vaccines

XX

PS Example 4; Page 124; 150pp; French.

XX

CC AAV03575-606 represent sequences that are present in *Neisseria*

CC meningitidis and *N. gonorrhoeae* but not in *N. lactamica*, except for the

CC genes involved in biosynthesis of the capsule polysaccharide, frrA or C,

CC opc, porA, rotamase, sequence icl106, Iga protease, pillin, pilC,

CC proteins which bind transferrin and opacity proteins. The DNA sequences

CC are responsible for the differences in pathogenicity between *N.*

CC meningitidis and *N. gonorrhoeae*, specifically they include the genes that

CC allow *N. meningitidis* to cross the blood-brain barrier. DNA sequences

CC common to *N. meningitidis* and *N. gonorrhoeae*, but absent from *N.*

CC lactamica, are responsible for colonisation and penetration of the

CC mucosa. The DNA sequences can be used to produce probes and primers, and

CC antibodies produced against the encoded proteins are used in standard

CC hybridisation/immunodassay processes for diagnosis of *N. meningitidis*

CC infection, particularly meningitis.

XX

SQ Sequence 270 BP; 72 A; 67 C; 42 G; 89 T; 0 other;

Query Match 100.0%; Score 270; DB 19; Length 270;

Best Local Similarity 100.0%; Pred. No. 3.5e-70;

Matches 270; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AATCTTCGCGACGGGAGGCTGTTTCTTCCTTCCTTCGACCGGATTCCTCAATA 60

DB 1 AATCTTCGCGACGGGAGGCTGTTTCTTCCTTCCTTCGACCGGATTCCTCAATA 60

QY 61 AAATCATTTGATTCATCGAAGTTCATTCCTATACCATTCCTTTAATACGATTTTATG 120

DB 61 AAATCATTTGATTCATCGAAGTTCATTCCTATACCATTCCTTTAATACGATTTTATG 120

QY 121 CTCGGTTTATCGAATACCTAACTTCCACTTCCTAGCAGCATGATGAGGATTCGC 180

DB 121 CTCGGTTTATCGAATACCTAACTTCCACTTCCTAGCAGCATGATGAGGATTCGC 180

QY 181 TATCACTCGGCAATCGCAGGACAGTGTGGAATACAACTTTACACCCCAATGTTTCA 240

DB 181 TATCACTCGGCAATCGCAGGACAGTGTGGAATACAACTTTACACCCCAATGTTTCA 240

QY 241 TTACGGTTGGCTCGAAACTCAATTTCAATT 270

DB 241 TTACGGTTGGCTCGAAACTCAATTTCAATT 270

RESULT 2

AAAL5393

ID AAAL5393 standard; DNA; 270 BP.

XX

AC AAAL5393;

XX

DT 04-SEP-2000 (first entry)

XX

DE Genomic fragment of *Neisseria meningitidis* Z2491.

XX

KW Pathogenic strain; *Neisseria*; vaccine; *Neisseria* infection; ss.

XX

OS *Neisseria meningitidis*.

XX

PN WO200026375-A2.

XX

PD 11-MAY-2000.

XX

PF 28-OCT-1999; 99WO-FR02643.

XX

PR 30-OCT-1998; 98FR-0013693.

XX

XX

(INMR ) PASTEUR MERIEUX SERUMS & VACCINS SA.  
(INRM ) INSERM INST NAT SANTE & RECH MEDICALE.  
Aujame L, Bouchardon A, Renauld-Mongenie G, Rokbi B, Nassif X;  
Tinsley C, Perrin A;  
XX WPI; 2000-365622/31.

XX

PT New polypeptide specific for pathogenic *Neisseria* useful in therapeutic  
or preventative vaccines and for diagnosis -  
XX

CL Claim 1; Page 42; 187pp; French.

XX

CC The present sequence represents a genomic fragment of *Neisseria*  
meningitidis. The specification describes proteins that are specific  
for pathogenic strains of *Neisseria*. The polynucleotides, polypeptides,  
or their antigenic fragments, are used in vaccines to treat or protect  
CC against *Neisseria* infections, particularly by *N. meningitidis*. The  
polynucleotide sequences are also used for recombinant production of  
CC the polypeptide and to produce attenuated *Neisseria* strains that  
overexpress it, or express it in a non-toxic mutant form.  
CC note: the present sequence is included in Claim 1, but it is those  
CC sequences that do not include the present sequence that are actually  
CC claimed.

XX

SQ Sequence 270 BP; 72 A; 67 C; 42 G; 89 T; 0 other;

Query Match 100.0%; Score 270; DB 21; Length 270;

Best Local Similarity 100.0%; Pred. No. 3.5e-70;

Matches 270; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AATCTTCGCGACGGGAGGCTGTTTCTTCCTTCCTTCGACCGGATTCCTCAATA 60

DB 1 AATCTTCGCGACGGGAGGCTGTTTCTTCCTTCCTTCGACCGGATTCCTCAATA 60

QY 61 AAATCATTTGATTCATCGAAGTTCATTCCTATACCATTCCTTTAATACGATTTTATG 120

DB 61 AAATCATTTGATTCATCGAAGTTCATTCCTATACCATTCCTTTAATACGATTTTATG 120

QY 121 CTCGGTTTATCGAATACCTAACTTCCACTTCCTAGCAGCATGATGAGGATTCGC 180

DB 121 CTCGGTTTATCGAATACCTAACTTCCACTTCCTAGCAGCATGATGAGGATTCGC 180

QY 181 TATCACTCGGCAATCGCAGGACAGTGTGGAATACAACTTTACACCCCAATGTTTCA 240

DB 181 TATCACTCGGCAATCGCAGGACAGTGTGGAATACAACTTTACACCCCAATGTTTCA 240

QY 241 TTACGGTTGGCTCGAAACTCAATTTCAATT 270

DB 241 TTACGGTTGGCTCGAAACTCAATTTCAATT 270

RESULT 3

ABZ37755/c

ID ABZ37755 standard; DNA; 855 BP.

XX

AC ABZ37755;

XX

DT 07-MAR-2003 (first entry)

XX

DE N. gonorrhoeae nucleotide sequence SEQ ID 99.

XX

KW Antibacterial; infection; vaccine; gene therapy; gene; ds.

XX

OS *Neisseria gonorrhoeae*.

XX

PN WO200279243-A2.

XX

PD 10-OCT-2002.

XX

PF 12-FEB-2002; 2002WO-IB02069.

XX

PR 12-FEB-2001; 2001GB-0003424.

```

XX PA (CHIR-) CHIRON SPA.
XX PI Fontana MR, Pizza M, Masignani V, Monaci E;
XX DR WPI; 2003-058415/05.
XX DR P-PSDB; ABP76785.
XX PT New protein from Neisseria gonorrhoeae, useful for the manufacture of a
XX DR medicament for treating or preventing N. gonorrhoeae infection -
XX PS Claim 6; Page 188; 815pp; English.
XX CC The present invention relates to proteins from Neisseria gonorrhoeae.
XX CC Also disclosed are the nucleic acid molecules encoding the proteins and
XX CC antibodies that specifically bind to the proteins. The composition
XX CC comprising the protein, nucleic acid or antibody is useful for the
XX CC manufacture of a medicament for treating or preventing N. gonorrhoeae
XX CC infection, this may be in the form of a vaccine or gene therapy.
XX CC Sequences given in records AB237706-AB242016 represent nucleic acid
XX CC molecules of the invention.
XX SQ Sequence 855 BP; 300 A; 151 C; 178 G; 226 T; 0 other;
Query Match 93.8%; Score 253.2; DB 25; Length 855;
Best Local Similarity 98.5%; Pred. No. 4.6e-65;
Matches 266; Conservative 0; Mismatches 3; Indels 1; Gaps 1;
QY 1 AATTCCTCCGACGGGAGGCTGTTTCTCCCTCTGTTCCGACCGATCTCAATA 60
Db 282 AATTCCTCCGACGGGAGGCTGTTTCTCCCTCTGTTCCGACCGATCTCAATA 223
QY 61 AAAATCATGATTTCATCGAAGTTCATTCCTATACCATTAATCTTTAATACGATTTATG 120
Db 222 AAAATCATGATTTCATCGAAGTTCATTCCTATGCAATTAATCTTTAATACGATTTATG 163
QY 121 CTCGGTTTATCGAATAACCTAACCTCCACTTCGTCAGACATGATCGTAGGCATTCGC 180
Db 162 CTCGGTTTATCGAATAACCTAACCTCCACTTCGTCAGACATGATCGTAGGCATTCGC 103
QY 181 TATCAACTCGGCAATCGCAGGAACAGTGTGCGAATACAAATCTTTACACCCAAATGTTCA 240
Db 102 TATCAACTCGGCAATCGCAGGAACAGTGTGCGAATAC-ATCTTTACACCCAAATGTTCA 44
QY 241 TTACGGTTGGCTCGAAATCAATTTCAATT 270
Db 43 TTACGGTTGGCTCGAAATCAATTTCAATT 14
RESULT 4
AAA15322/c
ID AAA15322 standard; DNA; 858 BP.
XX AC AAA15322;
XX DT 04-SEP-2000 (first entry)
XX DE DNA encoding a polypeptide of a Neisseria pathogenic strain.
XX KW Pathogenic strain; Neisseria; vaccine; Neisseria infection; ss.
XX OS Neisseria gonorrhoeae.
XX FH Key Location/Qualifiers
XX CDS 1...858
XX FT /*tag= a
XX PN WO200026375-A2.
XX PD 11-MAY-2000.
XX PF 28-OCT-1999; 99WO-FR02643.
XX PR
XX

```

```

PR 30-OCT-1998; 98FR-0013693.
XX (INMR ) PASTEUR MERIEUX SERUMS & VACCINS SA.
XX PA (INMR ) INSERM INST NAT SANTE & RECH MEDICALE.
XX PI
XX PI Aujame L, Bouchardon A, Renaud-Mongenie G, Rokbi B, Nassif X;
XX PI Tinsley C, Perrin A;
XX DR WPI; 2000-365622/31.
XX DR P-PSDB; AAY93292.
XX PT New polypeptide specific for pathogenic Neisseria useful in therapeutic
XX PT or preventative vaccines and for diagnosis -
XX PS Claim 4; Page 134-135; 187pp; French.
XX CC The present sequence encodes a protein that is specific for pathogenic
XX CC strains of Neisseria. The polynucleotides, polypeptides, or their
XX CC antigenic fragments, are used in vaccines to treat or protect against
XX CC Neisseria infections, particularly by N. meningitidis. The
XX CC polynucleotide sequence is also used for recombinant production of
XX CC the polypeptide and to produce attenuated Neisseria strains that
XX CC overexpress it, or express it in a non-toxic mutant form.
XX SQ Sequence 858 BP; 302 A; 151 C; 178 G; 227 T; 0 other;
Query Match 93.8%; Score 253.2; DB 21; Length 858;
Best Local Similarity 98.5%; Pred. No. 4.7e-65;
Matches 266; Conservative 0; Mismatches 3; Indels 1; Gaps 1;
QY 1 AATTCCTCCGACGGGAGGCTGTTTCTCCCTCTGTTCCGACCGATCTCAATA 60
Db 282 AATTCCTCCGACGGGAGGCTGTTTCTCCCTCTGTTCCGACCGATCTCAATA 223
QY 61 AAAATCATGATTTCATCGAAGTTCATTCCTATACCATTAATCTTTAATACGATTTATG 120
Db 222 AAAATCATGATTTCATCGAAGTTCATTCCTATGCAATTAATCTTTAATACGATTTATG 163
QY 121 CTCGGTTTATCGAATAACCTAACCTCCACTTCGTCAGACATGATCGTAGGCATTCGC 180
Db 162 CTCGGTTTATCGAATAACCTAACCTCCACTTCGTCAGACATGATCGTAGGCATTCGC 103
QY 181 TATCAACTCGGCAATCGCAGGAACAGTGTGCGAATACAAATCTTTACACCCAAATGTTCA 240
Db 102 TATCAACTCGGCAATCGCAGGAACAGTGTGCGAATAC-ATCTTTACACCCAAATGTTCA 44
QY 241 TTACGGTTGGCTCGAAATCAATTTCAATT 270
Db 43 TTACGGTTGGCTCGAAATCAATTTCAATT 14
RESULT 5
ABZ41353/c
ID ABZ41353 standard; DNA; 891 BP.
XX AC ABZ41353;
XX DT 07-MAR-2003 (first entry)
XX DE N. gonorrhoeae nucleotide sequence SEQ ID 7295.
XX KW Antibacterial; infection; vaccine; gene therapy; gene; ds.
XX OS Neisseria gonorrhoeae.
XX PN WO200279243-A2.
XX PD 10-OCT-2002.
XX PF 12-FEB-2002; 2002WO-IB02069.
XX PR 12-FEB-2001; 2001GB-0003424.
XX

```



XX (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.  
 PA (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.  
 XX (SMIK ) SMITHKLINE BEECHAM.  
 XX Nassif X, Tinsley C, Achtman M, Merker P, Ruelle J;  
 PI Vinals C;  
 XX WPI; 1998-110594/10.  
 XX  
 XX Genes present in *Neisseria meningitidis* but not other *Neisseria*  
 PT species - and related host cells, RNA, anti-sense sequences,  
 PT polypeptide(s) and antibodies, useful for diagnosing *Neisseria*  
 PT meningitidis infection and in protective vaccines  
 XX  
 XX Example 4; Page 118; 150pp; French.  
 XX  
 XX AAV03575-606 represent sequences that are present in *Neisseria*  
 CC meningitidis and *N. gonorrhoeae* but not in *N. lactamica*, except for the  
 CC genes involved in biosynthesis of the capsule polysaccharide, fpa or C,  
 CC opc, porA, rotamase, sequence IC1106, IGA protease, pillin, pilC,  
 CC proteins which bind transferrin and opacity proteins. The DNA sequences  
 CC are responsible for the differences in pathogenicity between *N.*  
 CC meningitidis and *N. gonorrhoeae*, specifically they include the genes that  
 CC allow *N. meningitidis* to cross the blood-brain barrier. DNA sequences  
 CC common to *N. meningitidis* and *N. gonorrhoeae*, but absent from *N.*  
 CC lactamica, are responsible for colonisation and penetration of the  
 CC mucosa. The DNA sequences can be used to produce probes and primers, and  
 CC antibodies produced against the encoded proteins are used in standard  
 CC hybridisation/immunoassay processes for diagnosis of *N. meningitidis*  
 CC infection, particularly meningitis.  
 XX  
 SQ Sequence 269 BP; 95 A; 41 C; 65 G; 68 T; 0 other;  
 Query Match 45.1%; Score 121.8; DB 19; Length 269;  
 Best Local Similarity 98.4%; Pred. No. 2.4e-26;  
 Matches 123; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 AATCTTCGCGACGGGGAGGCTGTTTTCTCCCTTCGTTCCGACCGATTCTCAATA 60  
 DB 125 AATCTTCGCGACGGGGAGGCTGTTTTCTCCCTTCGTTCCGACCGATTCTCAATA 66  
 QY 61 AAAATCATGATTTCATCGAGTTTCATTCCTATACCATATCTTTAATACGATTATTG 120  
 DB 65 AAAATCATGATTTCATCGAGTTTCATTCCTATACCATATCTTTAATACGATTATTG 6  
 QY 121 CTCGG 125  
 DB 5 CTCGG 1  
 RESULT 8  
 AAA15389/c  
 ID AAA15389 standard; DNA; 269 BP.  
 XX  
 AC AAA15389;  
 XX  
 DT 04-SEP-2000 (first entry)  
 XX  
 DE Genomic fragment of *Neisseria meningitidis* Z2491.  
 XX  
 KW Pathogenic strain; *Neisseria*; vaccine; *Neisseria* infection; ss.  
 XX  
 OS *Neisseria meningitidis*.  
 XX  
 FN WO200026375-A2.  
 XX  
 PD 11-MAY-2000.  
 XX  
 PF 28-OCT-1999; 99WO-FR02643.  
 XX  
 PR 30-OCT-1998; 98FR-0013693.  
 XX

PA (INMR ) PASTEUR MERIEUX SERUMS & VACCINS SA.  
 XX (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.  
 PI Aujame L, Bouchardon A, Renauld-Mongenie G, Rokbi B, Nassif X;  
 XX Tinsley C, Perrin A;  
 DR WPI; 2000-365622/31.  
 XX  
 XX New polypeptide specific for pathogenic *Neisseria* useful in therapeutic  
 PT or preventative vaccines and for diagnosis  
 XX  
 XX Claim 1; Page 40; 187pp; French.  
 XX  
 CC The present sequence represents a genomic fragment of *Neisseria*  
 CC meningitidis. The specificatio describes proteins that are specific  
 CC for pathogenic strains of *Neisseria*. The polynucleotides, polypeptides,  
 CC or their antigenic fragments, are used in vaccines to treat or protect  
 CC against *Neisseria* infections, particularly by *N. meningitidis*. The  
 CC polynucleotide sequences are also used for recombinant production of  
 CC the polypeptide and to produce attenuated *Neisseria* strains that  
 CC overexpress it, or express it in a non-toxic mutant form.  
 CC note: the present sequence is included in Claim 1, but it is those  
 CC sequences that do not include the present sequence that are actually  
 CC claimed.  
 XX  
 SQ Sequence 269 BP; 95 A; 41 C; 65 G; 68 T; 0 other;  
 Query Match 45.1%; Score 121.8; DB 21; Length 269;  
 Best Local Similarity 98.4%; Pred. No. 2.4e-26;  
 Matches 123; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 AATCTTCGCGACGGGGAGGCTGTTTTCTCCCTTCGTTCCGACCGATTCTCAATA 60  
 DB 125 AATCTTCGCGACGGGGAGGCTGTTTTCTCCCTTCGTTCCGACCGATTCTCAATA 66  
 QY 61 AAAATCATGATTTCATCGAGTTTCATTCCTATACCATATCTTTAATACGATTATTG 120  
 DB 65 AAAATCATGATTTCATCGAGTTTCATTCCTATACCATATCTTTAATACGATTATTG 6  
 QY 121 CTCGG 125  
 DB 5 CTCGG 1  
 RESULT 9  
 ABL13122/c  
 ID ABL13122 standard; cDNA; 4951 BP.  
 XX  
 AC ABL13122;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE *Drosophila melanogaster* expressed polynucleotide SEQ ID NO 33848.  
 XX  
 KW *Drosophila*; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical; gene; ss.  
 XX  
 OS *Drosophila melanogaster*.  
 XX  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 XX 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEXE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX  
 XX WPI; 2001-656860/75.

DR P-PSDB; ABB69019.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions -

XX

XX Claim 1; SEQ ID NO 33848; 21pp + Sequence Listing; English.

XX

XX The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

CC sequences (ABL01840-ABL16175) and the encoded proteins

CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX

SQ Sequence 4951 BP; 1376 A; 1121 C; 1065 G; 1389 T; 0 other;

Query Match 15.8%; Score 42.6; DB 23; Length 4951;

Best Local Similarity 61.1%; Pred. No. 0.016; Indels 0; Gaps 0;

Matches 69; Conservative 0; Mismatches 44;

QY 30 CTTCCCTTCTGTCGACCGATTCTCAATAAAATCATTCGATTCATCGAAGTTCATTC 89

DB 4057 CATGCACCTCTGTAACCGTATACCTACAAAATATATCATATATTAACAATTTATCT 3998

QY 90 CTATACCATTTATCTTTAATACGATTTTATCGTCGGTTTATCGAATAACCTA 142

DB 3997 CTTTACCATTATACTAAAAACGATGATGTCGGATACACTGCATCATT 3945

RESULT 10

ABL15458/c

ID ABL15458 standard; cDNA; 5395 BP.

XX

XX ABL15458;

XX

XX 26-MAR-2002 (first entry)

XX

XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 40856.

XX

XX Drosophila; developmental biology; cell signalling; insecticide;

XX pharmaceutical; gene; ss.

XX

XX Drosophila melanogaster.

XX

XX WO200171042-A2.

XX

XX 27-SEP-2001.

XX

XX 23-MAR-2001; 2001WO-US09231.

XX

XX 23-MAR-2000; 2000US-191637P.

XX

XX 11-JUL-2000; 2000US-0614150.

XX

XX (PEXE ) PE CORP NY.

XX

XX Venter JC, Adams M, Li PWD, Myers EW;

XX

XX WPI; 2001-656860/75.

XX

XX P-PSDB; ABB71355.

XX

XX New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions -

XX

XX Claim 1; SEQ ID NO 40856; 21pp + Sequence Listing; English.

XX

XX The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

CC sequences (ABL01840-ABL16175) and the encoded proteins

CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX

SQ Sequence 5395 BP; 1528 A; 1109 C; 1141 G; 1617 T; 0 other;

Query Match 15.8%; Score 42.6; DB 23; Length 5395;

Best Local Similarity 61.1%; Pred. No. 0.016; Indels 0; Gaps 0;

Matches 69; Conservative 0; Mismatches 44;

QY 30 CTTCCCTTCTGTCGACCGATTCTCAATAAAATCATTCGATTCATCGAAGTTCATTC 89

DB 536 CATGCACCTCTGTAACCGTATACCTACAAAATATATCATATATTAACAATTTATCT 477

QY 90 CTATACCATTTATCTTTAATACGATTTTATCGTCGGTTTATCGAATAACCTA 142

DB 476 CTTTACCATTATACTAAAAACGATGATGTCGGATACACTGCATCATT 424

RESULT 11

ABN80204/c

ID ABN80204 standard; DNA; 6068 BP.

XX

XX ABN80204;

XX

XX 15-JUL-2002 (first entry)

XX

XX Human chemically modified disease associated gene SEQ ID NO 221.

XX

XX Human; development; homeobox gene; HOX; diabetes; cancer; apoptosis;

XX heart disease; epilepsy; histone deacetylation; muscular dystrophy;

XX dwarfism; single nucleotide polymorphism; SNP; cytosine methylation;

XX antidiabetic; cytostatic; anticonvulsant; ds.

XX

XX Homo sapiens.

XX

XX Synthetic.

XX

XX WO200200927-A2.

XX

XX 03-JAN-2002.

XX

XX 02-JUL-2001; 2001WO-EP07536.

XX

XX 30-JUN-2000; 2000DE-1032529.

XX

XX 01-SEP-2000; 2000DE-1043826.

XX

XX (EPIG-) EPIGENOMICS AG.

XX

XX Olek A, Piepenbrock C, Berlin K;

XX

XX WPI; 2002-130908/17.

XX

XX Novel nucleic acid useful for diagnosis and therapy of diseases

XX associated with development genes such as diabetes, comprises a

XX sequence of a segment of chemically pretreated DNA of genes associated

XX with development -

XX

XX Claim 1; SEQ ID NO 221; 27pp; English.

XX

XX The invention relates to a nucleic acid (i) comprising a sequence at

CC least 18 bases in length of a segment of chemically pretreated DNA (ii)

CC of genes associated with development selected from 87 genes listed in

CC the specification such as ACCPN, ABPN, or AP1 and comprising one of 350

CC sequences (ABN7994-ABN8033) or their complements. The invention is

CC useful for the diagnosis or therapy of diseases associated with

CC development genes, in particular disease related to homeobox containing



CC genes (Hox), like diabetes, cancer, apoptosis related diseases, syndromes  
 CC associated with congenital heart disease, epilepsy, diseases related to  
 CC histone deacetylation, Currarino syndrome, diseases related with the  
 CC development of the brain and limb girdle muscular dystrophy and dwarfism.  
 CC Oligomers specific to each of the genes are useful for detecting the  
 CC methylation state of all CpG dinucleotides within the 350 sequences or  
 CC (II) and their complementary sequences, as primer oligonucleotides for  
 CC the amplification of the 350 sequences, (II) and/or their complements and  
 CC as oligomer probes for detecting the cytosine methylation state and/or  
 CC single nucleotide polymorphisms (SNPs).  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification but is based on sequence information supplied to Derwent by  
 CC the European Patent Office.

CC Sequence 6068 BP; 1819 A; 81 C; 1390 G; 2778 T; 0 other;

Query Match 14.1%; Score 38.2; DB 24; Length 6068;  
 Best Local Similarity 59.8%; Pred. No. 0.34;  
 Matches 64; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 56 AATAAATAATCATTCATTTCGAGTTCTTCTATACCATTTATTAATACGATT 115  
 DB 4187 AAAAAAATAATTAACATAATAAATCTTCTACATTTATTTATATACCATTA 4128  
 QY 116 TTATGCTCCGGTTTATCGAATAACCTAATTCCTCCACTTCGAGCACA 162  
 DB 4127 TTATCCTACCGCTTATCAAAACACCACTAAATCCTCACTAACACA 4081

RESULT 12  
 AAX91990/c  
 ID AAX91990 standard; DNA; 1230025 BP.  
 AC AAX91990;  
 XX 13-SEP-1999 (first entry)  
 DE Nucleotide sequence of the complete genome of Chlamydia pneumoniae.  
 KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;  
 KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;  
 KW vaccine; neutralising epitope; ss.  
 XX Chlamydia pneumoniae.

OS WO9927105-A2.  
 PN 03-JUN-1999.  
 XX 20-NOV-1998; 98WO-IB01890.  
 PF 04-NOV-1998; 98US-0107078.  
 PR 21-NOV-1997; 97FR-0014673.  
 XX (GEST ) GENSET.

PA Griffais R;  
 PI WPI; 1999-357842/30.  
 XX Genome sequence of Chlamydia pneumoniae  
 XX Claim 1; Page 291-611; 1912pp; English.

CC The present sequence represents the complete genome of Chlamydia  
 CC pneumoniae, and encodes proteins AAY34584-Y35879. C. pneumoniae causes  
 CC respiratory disease such as pneumonia and bronchitis and is thought  
 CC to be a contributing factor in heart disease, sarcoidosis, sinusitis,  
 CC purulent otitis media, erythema nodosum or pharyngitis. The polypeptides  
 CC encoded by the open reading frames of the C. pneumoniae genome (see  
 CC AAY34584-Y35879) can be used in immunogenic compositions as vaccines.  
 CC Vectors containing C. pneumoniae nucleotide sequences can also be  
 CC used as immunogenic compositions, especially where the vector directs

CC the expression of a neutralising epitope of C. pneumoniae.

XX Sequence 1230025 BP; 367213 A; 249833 C; 249013 G; 363589 T; 377 other;

Query Match 13.9%; Score 37.6; DB 20; Length 1230025;  
 Best Local Similarity 54.3%; Pred. No. 2.5;  
 Matches 76; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 70 GATTTCATCGAAGTTTCATTCCTATACCATTTATTAATACGATTTTATGCTCCGGTTT 129

DB 929463 GATGCTTTCGGCTCTGAAGCCACACCATTTCTGATAATGATTCGCGCTTGTCCTCC 929404

QY 130 ATCGAATAACCTAATTCACATTCCTAGCATGATCGTAGGCAATTCGCTATCAATC 189

DB 929403 TCTAAAGATTCGATTTCTATTCATCGCGCAGCATCTAAGGAATTCATCAATCAGTTC 929344

QY 190 GCGAATCGCAGNACAGTGT 209

DB 929343 TTTTAAACACAGAAACGGAGT 929324

RESULT 13  
 AAX86191  
 ID AAX86191 standard; DNA; 7160 BP.  
 XX AAX86191;  
 AC AAX86191;  
 XX 22-SEP-1999 (first entry)  
 DE Nucleotide sequence of the TONNEAU gene.

XX TONNEAU; cell division; cell elongation; eukaryotic cell; plant cell;  
 KW floral structure; male sterility; anther dehiscence; petal removal;  
 KW apetalous seed; female sterility; ornamental plant; ss.

OS Arabidopsis thaliana.

XX WO9935269-A2.

PN 15-JUL-1999.

PD 08-JAN-1999; 99WO-FR00021.

XX 09-JAN-1998; 98FR-0000162.

PR (INRG ) INRA INST NAT RECH AGRONOMIQUE.

PA Bouchez D, Nacry P;

PI WPI; 1999-430396/36.

XX P-PSDB; AAY23987, AAY23988.

DR Controlling division and elongation of eukaryotic plant cells

XX Example 2; Page 42-45; 48pp; French.

XX The present sequence represents the TONNEAU gene of Arabidopsis  
 CC thaliana. The specification describes a method to control division  
 CC and/or elongation of eukaryotic cells, and in particular plant  
 CC cells, by modifying the expression and/or activity of a TONNEAU  
 CC protein in the cells. The method is useful for the control of division  
 CC and/or elongation of plant cells. The TONNEAU protein coding  
 CC sequence can be inhibited by an antisense sequence or by mutagenesis.  
 CC The expression of TONNEAU-type genes can be tissue specific.  
 CC Alternatively, TONNEAU protein inhibitors are used to control its  
 CC activity. In particular, the method is used to alter floral structure  
 CC to induce male sterility (by reducing size and blocking dehiscence of  
 CC anthers), removal of petals (by producing an apetalous seed) or  
 CC inducing female sterility in ornamental species at the end of the  
 CC flowering period.

XX Sequence 7160 BP; 2086 A; 1287 C; 1349 G; 2438 T; 0 other;



```
FT CDS /note= "encodes AAY93248"
FT 10134..10430
FT /*tag= m
FT /product= "virulence protein"
FT /note= "encodes AAY93249"
FT 10459..10779
FT /*tag= n
FT /product= "virulence protein"
FT /note= "encodes AAY93250"
XX
PN WO200028038-A2.
XX
XX 18-MAY-2000.
XX
XX 09-NOV-1999; 99WO-GB03721.
XX
XX 09-NOV-1998; 98GB-0024569.
XX 09-NOV-1998; 98GB-0024570.
XX 17-DEC-1998; 98GB-0027814.
XX 17-DEC-1998; 98GB-0027815.
XX 17-DEC-1998; 98GB-0027816.
XX 17-DEC-1998; 98GB-0027818.
XX 13-JAN-1999; 99GB-0000708.
XX 13-JAN-1999; 99GB-0000710.
XX 13-JAN-1999; 99GB-0000711.
XX 28-JAN-1999; 99GB-0001915.
XX
XX (MICR-) MICROSCIENCE LTD.
XX
XX Crooke HR, Clarke EE, Everest PH, Dougan G, Holden DW, Shea JE;
PI Feldman RG;
XX
XX WPI; 2000-376550/32.
XX
XX P-PSDB; AAY93237, AAY93238, AAY93239, AAY93240, AAY93241, AAY93242,
XX AAY93243.
XX
XX Peptide encoded by an operon including genes from Escherichia coli for
PT screening potential drugs, detecting virulence and treating conditions
PT associated with infection by a Gram negative bacterium -
XX
XX Disclosure; Page 83-101; 122pp; English.
XX
XX The present sequence encodes Escherichia coli virulence proteins
CC The specification describes virulence proteins which are encoded
CC by an operon including tatA, tatB, tatC, tatE, mdoG, creC, recG, yggN,
CC eckl, iroD, iroC, iroE, mtd2 or msl-16 genes obtained from Escherichia-
CC coli K1. The virulence proteins and polynucleotides, and their vaccines
CC are useful for screening potential drugs, for the detection of virulence,
CC and for treating or preventing conditions associated with infection by
CC a Gram negative bacterium particularly Escherichia coli.
XX
XX Sequence 11165 BP; 3592 A; 2087 C; 2458 G; 3028 T; 0 other;
SQ
Query Match 13.3%; Score 36; DB 21; Length 11165;
Best Local Similarity 56.4%; Pred No. 1.8;
Matches 88; Conservative 0; Mismatches 65; Indels 3; Gaps 1;
QY 47 CCGATTCTCAATAAATATCATTTGATTTTCATCGAAGTTTCATTCCTATACCATTTATCTTTA 106
Db 5062 CCAATAGTCACACCACTCTCTCAATTCATCGGTTGTCATACCTATCCATCATCT-- 5006
QY 107 ATACGATTTTATGTCGCGTTTATCGAATTAACCTAATTCCTTCCTAGGACATGCA 166
Db 5005 CTCAGATAAGAAGATTTTCCTTCTCTAAATAAATCAACTTCGACATTATCAGCATAGGCA 4946
QY 167 TCGTAGGCAATTCGGTATCATCAACTCGGCAATCGCAGGA 202
Db 4945 TCATGAGCATTTTAAATAACTCACTCAAGGCAGTA 4910
```

Search completed: November 15, 2003, 00:35:37  
Job time : 178.258 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 14, 2003, 23:58:39 ; Search time 1801.19 Seconds  
(without alignments)  
3643.257 Million cell updates/sec

Title: US-09-928-457-88

Perfect score: 270

Sequence: 1 AATTCTCCGACGGGAGG.....CTCGAAACTCAATTCAATT 270

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:

1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rod:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	42.6	15.8	521	10 AW942932	AW942932 LD29090.3
2	41.2	15.3	548	9 AV955775	AV955775 AV955775
3	40	14.8	722	13 BU437178	BU437178 604146050
4	40	14.8	753	9 AJ450902	AJ450902 AJ450902

C	5	40	14.8	866	13	BU421944	BU421944
C	6	39.4	14.6	637	28	AZ524804	AZ524804
C	7	39	14.4	437	9	AJ456670	AJ456670
C	8	38.2	14.1	946	13	BQ735322	BQ735322
C	9	37.4	13.9	527	12	BM256945	BM256945
C	10	37.2	13.8	1201	13	BM256945	BM256945
C	11	37	13.7	280	29	BM256945	BM256945
C	12	37	13.7	1030	29	BM256945	BM256945
C	13	36.8	13.6	688	28	BM256945	BM256945
C	14	36.8	13.6	801	10	BE225447	BE225447
C	15	36.8	13.6	809	29	CC122282	CC122282
C	16	36.8	13.6	890	28	AZ551079	AZ551079
C	17	36.6	13.6	921	28	BH151796	BH151796
C	18	36.6	13.6	492	28	AA550645	AA550645
C	19	36.6	13.6	500	9	AU086225	AU086225
C	20	36.2	13.4	437	28	AZ173796	AZ173796
C	21	36.2	13.4	446	29	BZ417183	BZ417183
C	22	36.2	13.4	708	29	BZ653531	BZ653531
C	23	36.2	13.4	759	13	BQ704392	BQ704392
C	24	36.2	13.4	777	29	BZ391784	BZ391784
C	25	36.2	13.4	815	29	BZ637852	BZ637852
C	26	36.2	13.4	1201	13	BX385455	BX385455
C	27	36	13.3	441	9	AI485891	AI485891
C	28	36	13.3	542	13	BQ487518	BQ487518
C	29	35.8	13.3	863	28	BH574782	BH574782
C	30	35.8	13.3	1201	9	AL538100	AL538100
C	31	35.8	13.3	1201	13	BX423406	BX423406
C	32	35.8	13.3	1293	29	CC293132	CC293132
C	33	35.6	13.2	499	12	BJ108544	BJ108544
C	34	35.6	13.2	517	13	BZ37942	BZ37942
C	35	35.6	13.2	1104	13	BX36473	BX36473
C	36	35.6	13.2	1201	9	AL526873	AL526873
C	37	35.2	13.0	397	13	BX101702	BX101702
C	38	35.2	13.0	535	28	AQ512978	AQ512978
C	39	35.2	13.0	590	29	BZ421054	BZ421054
C	40	35.2	13.0	634	28	AZ523475	AZ523475
C	41	35.2	13.0	883	29	CNS035LS	CNS035LS
C	42	35.2	13.0	1584	29	CC237385	CC237385
C	43	35	13.0	400	11	AY067665	AY067665
C	44	35	13.0	770	29	CC131165	CC131165
C	45	35	13.0	864	28	AZ686202	AZ686202

#### ALIGNMENTS

RESULT 1  
AW942932  
LOCUS  
DEFINITION  
LD29090.3prime LD Drosophila melanogaster embryo pot2 Drosophila  
melanogaster cDNA clone LD29090 3, mRNA sequence.  
ACCSSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Drosophila melanogaster (fruit fly)  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE  
1 (bases 1 to 521)  
Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., St.C., Tsang, G.,  
Lewis, S. and Rubin, G.M.  
BDGP/HMI Drosophila EST Project  
Unpublished  
JOURNAL  
COMMENT  
Other ESTs: LD29090.5prime  
Contact: Stapleton, M.  
BDGP  
Lawrence Berkeley National Lab  
One Cyclotron Rd  
Berkeley, CA 94720, USA  
Fax: 510 486 6798  
Email: [http://www.fruitfly.org/EST\\_est@fruitfly.berkeley.edu](http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu)  
Based upon the presence of a XhoI site followed by a run of 14 or  
more T residues at the beginning of the sequence, this clone was

polyadenylated. The resulting Poly-T sequence has been removed. hit  
 Genomic AB003678; Drosophila melanogaster genomic scaffold  
 142000013386035 section 3 of 105, complete sequence.: 02/25/2001  
 Plate: LD.290 row: H column: 6  
 High quality sequence stop: 382.

## FEATURES

Location/Qualifiers  
 1..521

/organism="Drosophila melanogaster"

/mol\_type="mRNA"

/db\_xref="taxon:7227"

/clones="LD29030"

/sex="male and female"

/dev\_stage="0 to 24 hours mixed stage embryonic"

/lab\_host="XLI Blue"

/clone\_lib="LD Drosophila melanogaster embryo pot2"

/note="Organ: embryo; Vector: pOT2; Site 1: EcORI; Site 2:

XhoI; Sized fractionated cDNAs were directly ligated into

pOT2."

147 a 129 c 100 g 145 t

BASE COUNT

ORIGIN

Query Match 15.8%; Score 42.6; DB 10; Length 521;

Best Local Similarity 61.1%; Pred. No. 0.65;

Matches 69; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 30 CTTCCCTCTCTTCGACCGATTCTCAATAAAATCATTGATTCATCGAAGTTCAATC 89

DB 154 CATGCACCTCTGAACCGTATACCTACAAAATATATCATATATTAACAAATTTATCT 213

QY 90 CTATACCATTCATTTATACGATTTATCTCGGTTTATCGAATAACCTA 142

DB 214 CTTTACCATTATTAATAAAACGATGATCTCGATACACTGCATCACTTA 266

## RESULT 2

AV995775/c

LOCUS AV995775 548 bp mRNA linear EST 15-MAR-2002

DEFINITION AV995775 Nori Satoh unpublished cDNA library, tailbud embryo Ciona

intestinalis cDNA clone citb42e21 5', mRNA sequence.

ACCESSION AV995775

VERSION AV995775.1 GI:19487109

KEYWORDS EST.

SOURCE Ciona intestinalis

ORGANISM Ciona intestinalis

Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;

Phlebobranchia; Clonidae; Ciona.

REFERENCE 1 (bases 1 to 548)

Satch,N., Satou,Y., Kohara,Y. and Shin-i,T.

Expressed genes in Ciona intestinalis

Unpublished

Contact: Nori Satoh

Department of Zoology

Kyoto University

Sakyo-ku, Kyoto, Kyoto 606-8502, Japan

Tel: 81-75-753-4081

Fax: 81-75-705-1113

Email: satcho@ascidian.zool.kyoto-u.ac.jp.

Location/Qualifiers

1..548

/organism="Ciona intestinalis"

/mol\_type="mRNA"

/db\_xref="taxon:7719"

/clone="citb42e21"

/tissue\_type="whole animal"

/dev\_stage="tailbud embryo"

/clone\_lib="Nori Satoh unpublished cDNA library, tailbud

embryo"

167 a 95 c 138 g 148 t

BASE COUNT

ORIGIN

Query Match 15.3%; Score 41.2; DB 9; Length 548;

Best Local Similarity 57.9%; Pred. No. 1.6;

Matches 73; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 64 ATCATTTGATTTTCATCGAAGTTTCATCTTATACCATTTATCTTTAATAACGATTTATGCTC 123

DB 314 ATATTTCATTTCTCTGCTTTAAATTCACACCATTTATCCCGAATTCGATTTTCAAG 255

QY 124 CGGTTTTCGAATACCTTAACCTTCACCTTCGTCAGTACATCATCGTAGGATTCGGTAT 183

DB 254 CCCAAATCGACCAATTTGACTTGAATCGACGTTGCACCGCGTCCAGTGCATTTCTCAGT 195

QY 184 CAACCTC 189

DB 194 TAGTTC 189

RESULT 3

BU437178/c

LOCUS BU437178 722 bp mRNA linear EST 29-NOV-2002

DEFINITION 604146050F1 CSEQORBN11 Gallus gallus cDNA clone ChEST986c14 5', mRNA

sequence.

ACCESSION BU437178

VERSION BU437178.1 GI:25926489

KEYWORDS EST.

SOURCE Gallus gallus (chicken)

ORGANISM Gallus gallus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Archosauria; Aves; Neognathae; Galliformes; Phasianidae;

Phasianinae; Gallus.

REFERENCE 1 (bases 1 to 722)

Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,

Pong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.

A Comprehensive Collection of Chicken cDNAs

Curr. Biol. 12 (22), 1965-1969 (2002)

2235534

PUBMED

COMMENT

Contact: Simon Hubbard

Department of Biomolecular Sciences

University of Manchester Institute of Science and Technology (UMIST)

PO Box 88, Manchester, M60 1QD, UK

Tel: 01612008930

Fax: 01612360409

Email: Simon.Hubbard@umist.ac.uk.

Location/Qualifiers

1..722

/organism="Gallus gallus"

/mol\_type="mRNA"

/strains="Layer and broiler"

/db\_xref="taxon:9031"

/clone="ChEST986c14"

/sex="Male and female"

/tissue\_type="muscle"

/dev\_stage="adult"

/lab\_host="DH10B"

/clone\_lib="CSEQORBN11"

/note="Vector: pBluescript II KS(+); Site 1: EcORI;

Site 2: NotI; This normalized library was constructed from

1 million independent clones. cDNA synthesis was initiated

using an oligo(dT) primer, using methylated C in the first

strand synthesis reaction. Following this first strand

reaction, double-stranded cDNA was blunted, ligated to

NotI adapters, digested with EcORI, size-selected, and

cloned into the NotI and EcORI compatible sites of a

custom modified MCS of the pBluescript (KS+) vector. The

library was normalized in 2 rounds using conditions

adapted from Soares et al., PNAS (1994) 91: 9228-9232 and

Bonaldi et al., Genome Research 6 (1996): 791, except that

a significantly longer reannealing hybridization was

used."

BASE COUNT 224 a 139 c 166 g 192 t 1 others

ORIGIN

Query Match 14.8%; Score 40; DB 13; Length 722;

Best Local Similarity 54.9%; Pred. No. 3.4;

Matches 79; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 62 AATCATGATTCATCGAGTTCATCTCTATACCATTCATCTTTAATAACGATTTATGC 121  
 Db 171 ATAACAGGAACATCAGCGACCTTGATTCATTCGCAATGCTCTTACTTCTATTGTGCA 112  
 QY 122 TCCGGTTTATCGAATAACCTAACTCCACTTCGCTAGCACATGCATCGTAGCATTCGCT 181  
 Db 111 AACCGTAATTCAGTCTATGTCATAGTAGTGGCACTGGCATCCAGAGGATTTTCT 52  
 QY 182 ATCAACTCGGCAATCGCAGGAACA 205  
 Db 51 ATCAGCTCCTTCACCACGCTGACA 28

RESULT 4  
 AJ450902/c  
 LOCUS AJ450902 753 bp mRNA linear EST 19-APR-2002  
 DEFINITION AJ450902 riken1 Gallus gallus cdna clone 26p22r1, mRNA sequence.  
 ACCESSION AJ450902  
 VERSION AJ450902.1 GI:20218123  
 KEYWORDS EST.

SOURCE Gallus gallus (chicken)  
 ORGANISM Gallus gallus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
 Phasianinae; Gallus.  
 REFERENCE 1 (bases 1 to 753)  
 AUTHORS Buerstedde, J.M.  
 TITLE Gallus gallus bursal lymphocyte EST  
 JOURNAL Unpublished  
 COMMENT Contact: Buerstedde JM  
 Cellular Immunology  
 Heinrich-Pette-Institut  
 Martinistr. 52, 20251 Hamburg, Germany  
 Email: URL: http://genetics.hpi.uni-hamburg.de/dt40est.html.  
 FEATURES Location/Qualifiers  
 source 1..753  
 /organism="Gallus gallus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9031"  
 /clone="26p22r1"  
 /cell\_type="bursal lymphocyte"  
 /dev\_stage="2-3 weeks old"  
 /clone\_lib="riken1"  
 /note="CB inbred strain"

BASE COUNT 226 a 151 c 179 g 194 t 3 others

ORIGIN  
 Query Match 14.8%; Score 40; DB 9; Length 753;  
 Best Local Similarity 54.9%; Pred. No. 3.4;  
 Matches 79; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 62 AATCATGATTCATCGAGTTCATCTCTATACCATTCATCTTTAATAACGATTTATGC 121  
 Db 260 ATAACAGGAACATCAGCGACCTTGATTCATTCGCAATGCTCTTACTTCTATTGTGCA 201  
 QY 122 TCCGGTTTATCGAATAACCTAACTCCACTTCGCTAGCACATGCATCGTAGCATTCGCT 181  
 Db 200 AACCGTAATTCAGTCTATGTCATAGTAGTGGCACTGGCATCCAGAGGATTTTCT 141  
 QY 182 ATCAACTCGGCAATCGCAGGAACA 205  
 Db 140 ATCAGCTCCTTCACCACGCTGACA 117

RESULT 5  
 BU421944/c  
 LOCUS BU421944 866 bp mRNA linear EST 29-NOV-2002  
 DEFINITION 603955426F1 CSEQRBN09 Gallus gallus cdna clone CHES1922c8 5', mRNA  
 sequence.  
 ACCESSION BU421944  
 VERSION BU421944.1 GI:95914615

KEYWORDS EST.  
 SOURCE Gallus gallus (chicken)  
 ORGANISM Gallus gallus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
 Phasianinae; Gallus.  
 REFERENCE 1 (bases 1 to 866)  
 AUTHORS Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,  
 Pong, W.R., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.D.  
 TITLE A Comprehensive Collection of Chicken cDNAs  
 JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)  
 MEDLINE 22335534  
 PUBMED 12445392  
 COMMENT Contact: Simon Hubbard  
 Department of Biomolecular Sciences  
 University of Manchester Institute of Science and Technology (UMIST)

PO Box 88, Manchester, M60 1QD, UK  
 Tel: 01612008930  
 Fax: 01612360409  
 Email: Simon.Hubbard@umist.ac.uk.

FEATURES Location/Qualifiers  
 source 1..866  
 /organism="Gallus gallus"  
 /mol\_type="mRNA"  
 /strain="layer and broiler"  
 /db\_xref="taxon:9031"  
 /clone="ChEST922c8"  
 /sex="Male and female"  
 /tissue\_type="Chondrocytes isolated from growth plate cartilage"  
 /dev\_stage="adult"  
 /lab\_host="DH10B"  
 /clone\_lib="CSEQRBN09"  
 /note="Vector: pBluescript II KS(+); Site 1: EcoRI; Site 2: NotI; this normalized library was constructed from 1 million independent clones. cDNA synthesis was initiated using an oligo(dT) primer, using methylated C in the first strand synthesis reaction. Following this first strand reaction, double-stranded cDNA was blunted, ligated to NotI adapters, digested with EcoRI, size-selected, and cloned into the NotI and EcoRI compatible sites of a custom modified MCS of the pBluescript (KS+) vector. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6 (1996): 791, except that a significantly longer reannealing hybridization was used."

BASE COUNT 261 a 176 c 201 g 228 t

ORIGIN  
 Query Match 14.8%; Score 40; DB 13; Length 866;  
 Best Local Similarity 54.9%; Pred. No. 3.4;  
 Matches 79; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 62 AATCATGATTCATCGAGTTCATCTCTATACCATTCATCTTTAATAACGATTTATGC 121  
 Db 261 ATAACAGGAACATCAGCGACCTTGATTCATTCGCAATGCTCTTACTTCTATTGTGCA 202  
 QY 122 TCCGGTTTATCGAATAACCTAACTCCACTTCGCTAGCACATGCATCGTAGCATTCGCT 181  
 Db 201 AACCGTAATTCAGTCTATGTCATAGTAGTGGCACTGGCATCCAGAGGATTTTCT 142  
 QY 182 ATCAACTCGGCAATCGCAGGAACA 205  
 Db 141 ATCAGCTCCTTCACCACGCTGACA 118

RESULT 6  
 AZ524804  
 LOCUS AZ524804 637 bp DNA linear GSS 07-MAY-2001  
 DEFINITION 236PBE08 Pb MEN #21 Plasmodium berghei genomic 3', genomic survey  
 sequence.

ACCESSION AZ524804  
 VERSION AZ524804.1 GI:13965029  
 LOCUS GSS.  
 DEFINITION Plasmodium berghei  
 ACCESSION Plasmodium berghei  
 VERSION Plasmodium berghei  
 KEYWORDS Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 SOURCE 1 (bases 1 to 637)  
 ORGANISM Carlon, J.M.-R. and Dame, J.B.  
 REFERENCE The Plasmodium vivax and P. berghei gene sequence tag projects  
 AUTHORS Parasitol. Today (Regul. Ed.) 16 (10), 409 (2000)  
 TITLE The Plasmodium vivax and P. berghei gene sequence tag projects  
 JOURNAL Contact: Dame JB  
 COMMENT Dept. of Pathobiology, College of Veterinary Medicine  
 University of Florida  
 2015 SW 23rd Avenue, Bldg 1017, Gainesville, FL 32611, USA  
 Tel: 352 392 4700  
 Fax: 352 392 9704  
 Email: camej@mail.vetmed.ufl.edu  
 Seq primer: M13(-20) forward  
 Class: shotgun.  
 FEATURES  
 Location/Qualifiers  
 1..637  
 /organism="Plasmodium berghei"  
 /mol\_type="genomic DNA"  
 /strain="ANKA clone 15cyl (clone of the ANKA 8417 clone)"  
 /db\_xref="taxon:5821"  
 /dev\_stage="asexual blood forms"  
 /lab\_host="Mus musculus"  
 /clone\_lib="Pb MEN #21"  
 /note="Vector: pBluescript SK(+), phagemid excised from lambda ZAP; Site 1: EcoRV; Site 2: EcoRV; Genomic DNA was prepared from asynchronous blood stage forms of the cloned ANKA isolate of P. berghei grown in laboratory Swiss white mice. The DNA was purified from contaminating host DNA by Hoechst Dye 33258-CSCI ultracentrifugation and precipitated. Purified DNA was digested with Mung bean nuclease in the presence of 36-38% formamide at 50 C, as described (Vernick, K.D., Imberski, R.B., and McCutchan, T.F. 1988. Nucleic Acids Research 16:6883-6896). The ends of the digestion fragments were polished using T4 DNA polymerase, and the fragments size selected in the range 500-2000 bp. These were ligated into the EcoRV-cleaved and dephosphorylated pBluescript SK(+) vector. Recombinant plasmids were used to transform E. coli XL10-Gold host cells."  
 BASE COUNT 190 a 97 c 51 g 299 t  
 ORIGIN  
 Query Match 14.6%; Score 39.4; DB 28; Length 637;  
 Best Local Similarity 47.4%; Pred. No. 4.8;  
 Matches 118; Conservative 0; Mismatches 131; Indels 0; Gaps 0;  
 QY 18 AGGCTGTTTCTTCCCTTCGTGTCGACCGCATCTCAATAAATCAATGATTTTCAT 77  
 Db 316 AGTCATTTTGTGTACTTCCACCGCATATATTTTATCTCATAATATATAAATCT 375  
 QY 78 CGAGTTCATCCATACCATATCTTTTAATACGATTTTATGCTCCGTTTATCGAATA 137  
 Db 376 TATATTCGAATTTTTCAGTATTTTATTTTCATATTTTCATCTAATTCATTTAATA 435  
 QY 138 ACCTAACTTCCACTTCCGTAGACATGTCATCGTAGGATTCGCTATCAATCGGCAATCG 197  
 Db 436 TGTATTTTGTCTTAATCATTTTCAATACTCTTTTAAATTTTCAAAAACATCAAAATTTG 495  
 QY 198 CAGGACAGTGGCAATACATCTTTTACCCCAATGTCGATACGGTGGCTCGAA 257  
 Db 436 AAACATCATCTTCAATATTAATGTCATTTTAAATTCATTTTCTGTGGCATCTT 555  
 QY 258 CTCATTTTC 266  
 Db 556 TTCATATTC 564  
 RESULT 7

AJ456670/c  
 LOCUS 437 bp mRNA linear EST 22-APR-2002  
 DEFINITION riken1 Gallus gallus CDNA clone 9d22r1, mRNA sequence.  
 ACCESSION AJ456670  
 VERSION AJ456670.1 GI:20266766  
 KEYWORDS EST.  
 SOURCE Gallus gallus (chicken)  
 ORGANISM Gallus gallus  
 REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 AUTHORS Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
 TITLE Phasianinae; Gallus.  
 JOURNAL 1 (bases 1 to 437)  
 COMMENT Gallus gallus bursal lymphocyte EST  
 Unpublished  
 Contact: Buerstedde JM  
 Cellular Immunology  
 Heinrich-Pette-Institute  
 Martinistr. 52, 20251 Hamburg, Germany  
 Email: URL: http://genetics.hpi.uni-hamburg.de/dt40est.html.  
 FEATURES  
 Location/Qualifiers  
 1..437  
 /organism="Gallus gallus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9031"  
 /clone="9d22r1"  
 /cell\_type="bursal lymphocyte"  
 /dev\_stage="2-3 weeks old"  
 /clone\_lib="riken1"  
 /note="WCB inbred strain"  
 BASE COUNT 108 a 99 c 130 g 98 t 2 others  
 ORIGIN  
 Query Match 14.4%; Score 39; DB 9; Length 437;  
 Best Local Similarity 54.2%; Pred. No. 5.9;  
 Matches 78; Conservative 0; Mismatches 66; Indels 0; Gaps 0;  
 QY 62 AAATCATTTTCATCGAATTCATTCCTATACCATATCTTTAATACGATTTATGC 121  
 Db 336 ATACAGGAAATCATGACGACCTTGATTCATTCATTCATTTTGTCA 277  
 QY 122 TCGGTTTATCGAATAACCTTCCACTTCCTAGCAGCATGCTAGCATTCCT 181  
 Db 276 AACCGTAAATTCCTCAGTCTATGTCATAGTAGTGGCACTGGCATCAAGGATTTCT 217  
 QY 182 ATCAACTCGGCAATCGAGGAACA 205  
 Db 216 ATCAGTCTCTTCCACCACGCTGACA 193  
 RESULT 8  
 BQ735322/c  
 LOCUS 946 bp mRNA linear EST 16-JUL-2002  
 DEFINITION AGENCOURT\_8510714\_NICHD\_XGC\_Emb4\_Xenopus laevis cDNA clone  
 IMAGE:4684253 5', mRNA sequence.  
 ACCESSION BQ735322  
 VERSION BQ735322.1 GI:21874219  
 KEYWORDS EST.  
 SOURCE Xenopus laevis (African clawed frog)  
 ORGANISM Xenopus laevis  
 REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 AUTHORS Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae;  
 TITLE Xenopodinae; Xenopus.  
 JOURNAL 1 (bases 1 to 946)  
 COMMENT NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgabbs-r@mail.nih.gov  
 Tissue procurement: Dr. Igor Dawid  
 CDNA library preparation: Life Technologies, Inc.  
 CDNA library arrayed by: The I.M.A.G.E. Consortium (LLNL)



DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: NCI-CEAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.lnl.gov/bbrp/image/image.html](http://www-bio.lnl.gov/bbrp/image/image.html)  
 Plate: LLAW10568 row: c column: 06  
 High quality sequence stop: 637.

FEATURES  
source

```

1. 946
location/Qualifiers
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="IMAGE:4684253"
/dev_stage="embryo, stage 31-32"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NICHD XGC-Emb4"
/note="Organ: whole embryo; Vector: pCWV-
Nori; Site: 2; Salt; Cloned unidirectional
ct. Average insert size 2.1 kb. Constructed
Technologies. Note: This is a Xenopus Gen
Library "
```

BASE COUNT	300 a	187 c	199 g	259 e	1 others
ORIGIN					

Query Match	14.1%	Score 38.2;	DB 13;	Length 946;
Best Local Similarity	54.7%	Pred. No. 10;		
Matches 76; Conservative	0;	Mismatches 63;	Indels 0;	Gaps 0;

Qy 56 CATTGATTTCAGAGTTTCATTCTATACCATATCTTTAATAACGATTTTATGCTCG 125

Db 286 CAGAGGTTTCGTCATGTTTGATACCTTTCCATTGTCGGAACCTTCATTTGTGCAATC 237

Qy	126	GTATTTCGAATAACCTTA	CTACTCCACTTCCG	TAGCACATGCATCGTAGGCATTCGCTATCA	185
Db	226	CAAAATTTTCCAACTTA	ATTTTCAATGCTGGT	TGCACCTAGCATCTAAGGCATTTTCTACTA	167

Qy	186	ACTCGGCAATCGAGGAA	204
Db	166	GTTCCTTACCACACTAAC	148

RESULT 9  
BM256945  
LOCUS  
DEFINITION  
BM255945 527 bp mRNA linear EST 17-DEC-2001  
520410 MARCH Bos taurus cDNA 5', mRNA sequence.

ACCESSION BM256945  
VERSION BM256945.1  
KEYWORDS EST.  
SOURCE Bos taurus (cow)  
GI:1789254

ORGANISM      Bos taurus      Eukaryota      Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.

REFERENCE  
1 (bases 1 to 527)  
AUTHORS  
Smith, T. P. L., Grose, W. M., Erekling, B. A., Roberts, A. J., Stone, R. T., Casas, E., Wray, J. E., White, J., Cho, J., Fahrenkrug, S. C., Bennett, G. L., Heaton, M. P., Laegreid, W. W., Rohrer, G. A., Chicks-McKown, C. G., Berteau, G., Holt, I., Karamycheva, S., Liang, F., Quackenbush, J. and Keele, J. W.

**TITLE** Sequence evaluation of four pooled-tissue normalized bovine CDNA libraries and construction of a gene index for cattle  
**JOURNAL** Genome Res. 11 (4), 626-630 (2001)

21180013  
MEDLINE  
PUBMED  
COMMENT  
Contact: Smith TPL  
USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA

Tel: 402 762 4366  
Fax: 402 762 4390  
Email: smith@email.marc.usda.gov  
Single pass sequencing. Bases called and alt trimmed with phred  
v.0.980904.e. Vector identified by cross\_match with the -mnscore 18  
and -mismatch 12 options.  
pcp primers

FORWARD: AGGAACAGCTATGACCAT  
 BACKWARD: GTTTCCTCAGTCAGCAGC  
 Plate: 122 row: K column: 3  
 Seq primer: ATTAGTGTGACACTATAG.  
 location/Qualifiers  
 1. .527  
 FEATURES  
 source

```

/organism="Bos taurus"
/mol type="mrna"
/db xref="taxon:9913"
/tissue type="pooled"
/lab host="PHIOB"
/clont lib="MARC 3BOV"
/total=Vector: PCMV SPORT6;
Library made from pooled ti
macrophage, ovary, fetal se
longissimus muscle."
170 a      74 c    88 g   195 t
BASE COUNT

```

Query Match 13.9%; .Score 37.4; DB 12; Length 527;  
Best Local Similarity 54.8%; Pred. No. 16;  
Matches 74; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 12 ACGGGAGGCTGTTTTTCTCCCTCTGTTCCGA CGGATCTCAATAAAATCATGA 71

Db 283 ACGGTCAGACTAGGGTGTGCTGCTTATCTGTCGATTTTCAAACTGAAAGAGATGT 342

Qy	72	TTTCATCGAAGTTCATTCCTATACCATATCTTTTAATAACGATTTTATGCTCCGGTTTAT	131
Db	343	TTTATCTTTAGTTATCTCATTTATATTAATCTAAAGTAGATTTTATATAACAGATTTTAA	402

Qy	132	CGATAACCTA	146
Dd	403	TGCAAAAGTGTATTT	417

RESULT 10  
BX443799/c

LOCUS	DEFINITION	Accession	Size	Organism	Library	EST ID
LOCUS BX443739	CSODM001YA02	5-PRIME, MRNA	1201 bp	FETAL LIVER Homo sapiens	linear	EST 15-MAY-2003

ACCESSION	EX443799
VERSION	EX443799.
KEYWORDS	EST.
SOURCE	Homo sapi
ORGANISM	Homo sapi

REFERENCE  
1 (bases 1 to 1201)

AUTHORS	TITLE	JOURNAL	COMMENT

Genoscope - Centre National de Séquençage  
BP 191 91006 EVRY cedex - France  
Email: [seqref@genoscope.cns.fr](mailto:seqref@genoscope.cns.fr), Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
Library was constructed by Life Technology  
Invitrogen. This sequence belongs to sequen-  
more information about this cluster, see  
[http://www.genoscope.cns.fr/](http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CSODM001BA01QPI&c)  
<http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CSODM001BA01QPI&c>  
Feng Liang Email : [filang@lifetech.com](mailto:filang@lifetech.com) UR  
<http://fulllength.invitrogen.com/> Invitro-  
Paradise Avenue Genoscope sequene ID : CS

FEATURES	SOURCE
1. <i>Age</i>	1. <i>Age</i>
2. <i>Age</i>	2. <i>Age</i>
3. <i>Age</i>	3. <i>Age</i>
4. <i>Age</i>	4. <i>Age</i>
5. <i>Age</i>	5. <i>Age</i>
6. <i>Age</i>	6. <i>Age</i>
7. <i>Age</i>	7. <i>Age</i>
8. <i>Age</i>	8. <i>Age</i>
9. <i>Age</i>	9. <i>Age</i>
10. <i>Age</i>	10. <i>Age</i>
11. <i>Age</i>	11. <i>Age</i>
12. <i>Age</i>	12. <i>Age</i>
13. <i>Age</i>	13. <i>Age</i>
14. <i>Age</i>	14. <i>Age</i>
15. <i>Age</i>	15. <i>Age</i>
16. <i>Age</i>	16. <i>Age</i>
17. <i>Age</i>	17. <i>Age</i>
18. <i>Age</i>	18. <i>Age</i>
19. <i>Age</i>	19. <i>Age</i>
20. <i>Age</i>	20. <i>Age</i>
21. <i>Age</i>	21. <i>Age</i>
22. <i>Age</i>	22. <i>Age</i>
23. <i>Age</i>	23. <i>Age</i>
24. <i>Age</i>	24. <i>Age</i>
25. <i>Age</i>	25. <i>Age</i>
26. <i>Age</i>	26. <i>Age</i>
27. <i>Age</i>	27. <i>Age</i>
28. <i>Age</i>	28. <i>Age</i>
29. <i>Age</i>	29. <i>Age</i>
30. <i>Age</i>	30. <i>Age</i>
31. <i>Age</i>	31. <i>Age</i>
32. <i>Age</i>	32. <i>Age</i>
33. <i>Age</i>	33. <i>Age</i>
34. <i>Age</i>	34. <i>Age</i>
35. <i>Age</i>	35. <i>Age</i>
36. <i>Age</i>	36. <i>Age</i>
37. <i>Age</i>	37. <i>Age</i>
38. <i>Age</i>	38. <i>Age</i>
39. <i>Age</i>	39. <i>Age</i>
40. <i>Age</i>	40. <i>Age</i>
41. <i>Age</i>	41. <i>Age</i>
42. <i>Age</i>	42. <i>Age</i>
43. <i>Age</i>	43. <i>Age</i>
44. <i>Age</i>	44. <i>Age</i>
45. <i>Age</i>	45. <i>Age</i>
46. <i>Age</i>	46. <i>Age</i>
47. <i>Age</i>	47. <i>Age</i>
48. <i>Age</i>	48. <i>Age</i>
49. <i>Age</i>	49. <i>Age</i>
50. <i>Age</i>	50. <i>Age</i>
51. <i>Age</i>	51. <i>Age</i>
52. <i>Age</i>	52. <i>Age</i>
53. <i>Age</i>	53. <i>Age</i>
54. <i>Age</i>	54. <i>Age</i>
55. <i>Age</i>	55. <i>Age</i>
56. <i>Age</i>	56. <i>Age</i>
57. <i>Age</i>	57. <i>Age</i>
58. <i>Age</i>	58. <i>Age</i>
59. <i>Age</i>	59. <i>Age</i>
60. <i>Age</i>	60. <i>Age</i>
61. <i>Age</i>	61. <i>Age</i>
62. <i>Age</i>	62. <i>Age</i>
63. <i>Age</i>	63. <i>Age</i>
64. <i>Age</i>	64. <i>Age</i>
65. <i>Age</i>	65. <i>Age</i>
66. <i>Age</i>	66. <i>Age</i>
67. <i>Age</i>	67. <i>Age</i>
68. <i>Age</i>	68. <i>Age</i>
69. <i>Age</i>	69. <i>Age</i>
70. <i>Age</i>	70. <i>Age</i>
71. <i>Age</i>	71. <i>Age</i>
72. <i>Age</i>	72. <i>Age</i>
73. <i>Age</i>	73. <i>Age</i>
74. <i>Age</i>	74. <i>Age</i>
75. <i>Age</i>	75. <i>Age</i>
76. <i>Age</i>	76. <i>Age</i>
77. <i>Age</i>	77. <i>Age</i>
78. <i>Age</i>	78. <i>Age</i>
79. <i>Age</i>	79. <i>Age</i>
80. <i>Age</i>	80. <i>Age</i>
81. <i>Age</i>	81. <i>Age</i>
82. <i>Age</i>	82. <i>Age</i>
83. <i>Age</i>	83. <i>Age</i>
84. <i>Age</i>	84. <i>Age</i>
85. <i>Age</i>	85. <i>Age</i>
86. <i>Age</i>	86. <i>Age</i>
87. <i>Age</i>	87. <i>Age</i>
88. <i>Age</i>	88. <i>Age</i>
89. <i>Age</i>	89. <i>Age</i>
90. <i>Age</i>	90. <i>Age</i>
91. <i>Age</i>	91. <i>Age</i>
92. <i>Age</i>	92. <i>Age</i>
93. <i>Age</i>	93. <i>Age</i>
94. <i>Age</i>	94. <i>Age</i>
95. <i>Age</i>	95. <i>Age</i>
96. <i>Age</i>	96. <i>Age</i>
97. <i>Age</i>	97. <i>Age</i>
98. <i>Age</i>	98. <i>Age</i>
99. <i>Age</i>	99. <i>Age</i>
100. <i>Age</i>	100. <i>Age</i>

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CG00001YA02"
/tissue_type="FETAL LIVER"
/dev_stage="fetal"
/clone_lib="Homo sapiens FETAL LIVER"

```

/note="Organ: liver; Vector: pCMVSPORT.6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

BASE COUNT 268 a 219 c 364 g 188 t 162 others

ORIGIN

Query Match 13.8%; Score 37.2; DB 13; Length 1201;

Best Local Similarity 32.0%; Pred. No. 20; Mismatches 62; Conservative 46; Mismatches 86; Indels 0; Gaps 0;

QY 3 TTCTCCGACGGGAGCTGTTTTCTTCCCTTCTGTCGACCGATCTCAATAAA 62

DB 1123 TTYKCTGCTCCRMCCSGSCCVCCSTTTCCCYCAWTTTGGKMMHGWYCCCCCMC 1064

QY 63 AATCATTCATTTCATCGAAGTTCATCTCTATACCATTATCTTTAATAACGATTTATGCT 122

DB 1063 AAKKAASCTTYCTTGAATTCCTTTTACCTGACGTCVAAAAMWKKITTYT 1004

QY 123 CGGTTTATCGAATAACCTAACTTCCATCCGTCGACATGTCATGAGCATTCGCTA 182

DB 1003 CCRCTCCYCYTTTMMWMTTTCCKWTCOMYTGSMTCCTMMCCYCRADRTTSTTT 944

QY 183 TCACTCGGCAATC 196

DB 943 CMWRKKTCAKATC 930

RESULT 11

LOCUS BX293327

DEFINITION Arabidopsis thaliana T-DNA flanking sequence GK-477E04-019250, genomic survey sequence.

ACCESSION BX293327

VERSION BX293327.1 GI:28892323

KEYWORDS GSS.

SOURCE Arabidopsis thaliana (thale cress)

ORGANISM Arabidopsis thaliana

REFERENCE 1 Strizhov N., Li Y., Rosso M., Viehoveer P., Dekker K., Saedler H. and Weisshaar B.

A pipeline for automated high-throughput generation of PSTs (flanking sequence tags) from Arabidopsis thaliana T-DNA transformed lines

Unpublished

JOURNAL

REFERENCE 2

AUTHORS Rosso M., Strizhov N., Li Y., Reiss B., Dekker K. and Weisshaar B.

A new Arabidopsis thaliana T-DNA mutagenised population (GABI-Kat) for flanking sequence tag based reverse Genetics

Unpublished

JOURNAL

REFERENCE 3

AUTHORS Strizhov N., Rosso M., Li Y. and Weisshaar B.

Direct Submission

JOURNAL

TITLE Submitted (07-MAR-2003) Weisshaar B., Max-planck-Institut fuer

Zuechtungsforschung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany

This sequence is recovered from the left border of the T-DNA. It

indicates an insertion within the locus defined by clone F28P10.

The sequences are generated at the MPI for Plant Breeding Research

in the context of the GABI-Kat project. GABI-Kat is part of the

German Plant Genomics program designated 'GABI'. Information on

line availability can be found at:

http://www.mpiz-koeln.mpg.de/GABI-Kat/.

Location/Qualifiers

1..280

/organism="Arabidopsis thaliana"

/mol\_type="genomic DNA"

/strains="Columbia 0"

/db\_xref="taxon:3702"

/clone="GK-477E04-019250"

FEATURES

source

1..280

/organism="Arabidopsis thaliana"

/mol\_type="genomic DNA"

/strains="Columbia 0"

/db\_xref="taxon:3702"

/clone="GK-477E04-019250"

/clone\_lib="Arabidopsis thaliana T-DNA insertion lines"

/note="PCR was performed on DNA from Arabidopsis thaliana plants (T1) which were transformed with the T-DNA from vector pAC161. The lines contain one or more T-DNA insertions. The DNA fragment(s) resulting from the PCR were directly sequenced to determine the genomic sequence flanking the insertion. Sequences displaying significant similarity to the A. thaliana nuclear genome sequence were processed for submission. T-DNA derived sequences were removed"

BASE COUNT 74 a 56 c 44 g 105 t 1 others

ORIGIN

Query Match 13.7%; Score 37; DB 29; Length 280;

Best Local Similarity 58.7%; Pred. No. 19; Mismatches 64; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 28 TTCTTCCCTTCTGTTCCGACGATTCCTCAATAAAATCATTTATCCAAAGTTTCAAT 87

DB 90 TTCTTCTTATCCGATTCATCCCAITGACAAACCCCAAGCGTTGGTTATTCAATGCTCAT 149

QY 88 TCTATACATTTATCTTTAATAACGATTTTATGCTCCGGTTTATCGAT 136

DB 150 TCGTAACACGATATCCTCATTAGCTTTTCTTCCCATTTTGGCTTGAAT 198

RESULT 12

LOCUS CC254763

DEFINITION

CC254763

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1..1030

/organism="Callus gallus"

/mol\_type="genomic DNA"

/strains="Red Jungle Fowl"

/db\_xref="taxon:9031"

/clone="CH261-136L23"

/sex="female"

/cell\_line="UCD001, inbred 256"

/clone\_lib="CH261"

/note="Vector: pTABAC2.1; Site 1: EcoRI; Site 2: EcoRI; CH261 Female Chicken library - for library and clone ordering information: http://www.chori.org/bacpac"

BASE COUNT 345 a 138 c 145 g 400 t

ORIGIN

Query Match 13.7%; Score 37; DB 29; Length 1030;

Best Local Similarity 48.4%; Pred. No. 22; Mismatches 103; Conservative 0; Mismatches 110; Indels 0; Gaps 0;

```

QY 57 AATAAAATCATGATTCATCGAAGTTCATTCCTATACCAATATCTTTTAATAAGATT 116
DB 606 AATAAAACACAGTGTGTTGGTGGTATATTTTTCCTTTTTCCTTTTTCACATT 665
QY 117 TAGCTCGGTTTATCGAATACCTTAACCTTCACCTTCGTAGCAGATCGTAGGCAT 176
DB 666 TTTCCTTAGCTTTAGAAATTCCTTGCAGTCTCTTACAAATACAAAGTTAATACCGTTAA 725
QY 177 TCGCTATCAACTCGCAATCGCAGGACAGAGTGTGCGAATACAAATCTTTACACCCAAATGT 236
DB 726 GTGAAATTAAGTACCAATCACTTCTGTTTTCAGTACAGCTTTTAAAGCAAGTCT 785
QY 237 TCGATTACGGTTCGTCGAACTCAATTCAT 269
DB 786 TAAAGGAGTGTAGTCTCAATCAATTTTGAAT 818

RESULT 13
LOCUS BH934691
DEFINITION odel8d03.b1 B.oleracea002 Brassica oleracea genomic, genomic survey
sequence.
ACCESSION BH934691
VERSION BH934691.1 GI:23414757
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; rosids
; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 688)
AUTHORS Delehaanty,K., Fewell,G., Fulton,L., McCombie,W.R., Miner,T., Nash
,W., Rabinowicz,P.D. and Wilson,R.K.
TITLE Whole genome shotgun reads from Brassica oleracea
JOURNAL Unpublished
COMMENT Contact: Richard K. Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Plate: odel8 row: d column: 03
Seq primer: -21UpPOT forward
Class: shotgun
High quality sequence start: 8
High quality sequence stop: 551.
FEATURES
Source
Location/Qualifiers
1..688
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/db_xref="taxon:3712"
/clone_lib="B.oleracea002"
/note="Vector: pOTw13; Whole genome shotgun library from
flowering buds. DNA was purified from a crude nuclear
prep using Brassica oleracea TOL000DH3 buds provided by
Thomas Osborn at the University of Wisconsin. Genomic
DNA was provided by Pablo Rabinowicz (CSHL) and the
shotgun library prepared at Washington University Genome
Sequencing Center."
BASE COUNT 246 a 77 c 69 g 296 t
ORIGIN
Query Match 13.6%; Score 36.8; DB 28; Length 688;
Best Local Similarity 46.8%; Pred. No. 24;
Matches 116; Conservative 0; Mismatches 132; Indels 0; Gaps 0;

QY 21 CTTCCTTTTCTCCCTTCTGTTCGACCGATTCTCAATTAATAATCATTCATTTCATCGA 80
DB 404 CTTTTATCTTACCACCATTAAGAGACATTTTCAAAATACATTTTTCATTAAGTGA 463
QY 81 AGTTCATCTCATACCAATATCTTTAATAACGATTTTATGCTCGGTTTATCGAATAACC 140
DB 464 AAAAGACTCTTATACCAATGTATCTCTATATATAATAAATTTATTTAATAATAAA 523
QY 141 TAACTTCCATCTCGTAGACATGCATCGTAGGCATTCGCTATCAACTCGGCAATCGAG 200

```

```

DB 524 TAACACATAATAAAAAATTAATGTTTCGAATATACITTTTCAAAATCGAACTTTTC 583
QY 201 GAACAGTGTGCGAATCAATCTTTTACCCAAATGTTGATTCAGTTCGCTCGAAACTC 260
DB 584 ATAAATGTTTATTAATGAATTTTTCGAATGTTTATTTTTCATTTTTCAAAATTC 643
QY 261 AATTTCAA 268
DB 644 TTTTGGAA 651

RESULT 14
LOCUS BE225447/c
DEFINITION MD0355 Meloidegynie incognita J2 (#MD99-1) Meloidegynie incognita
cDNA clone 3108 5' similar to unknown Meloidegynie incognita cDNA,
mRNA sequence.
ACCESSION BE225447
VERSION BE225447.1 GI:8930683
KEYWORDS EST.
SOURCE Meloidegynie incognita (southern root-knot nematode)
ORGANISM Meloidegynie incognita
Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
Tylenchoidea; Heteroderidae; Meloidegyninae; Meloidegynie.
REFERENCE 1 (bases 1 to 801)
AUTHORS Dautova,M., Gommers,F., Bakker,J. and Smart,G.
TITLE 5' end expressed sequence tags from Meloidegynie incognita
JOURNAL Unpublished (2000)
COMMENT Contact: Smart G / Dautova M
Laboratory of Nematology
Wageningen University and Research Center
Box 8123, Binnenhaven 10, 6700 PD Wageningen, The Netherlands
Tel: 31 317 485 254
Fax: 31 317 484 254
Email: Geert.Smart@medew.nema.wau.nl,
Makedonka.Dautova@medew.nema.wau.nl
Insert Length: 801 Std Error: 0.00
Seq primer: T7 promoter primer
High quality sequence stop: 801.
FEATURES
Source
Location/Qualifiers
1..801
/organism="Meloidegynie incognita"
/mol_type="mRNA"
/db_xref="taxon:6306"
/clone="3108"
/dev_stage="second stage preparasitic juveniles (J2)"
/clone_lib="Meloidegynie incognita J2 (#MD99-1)"
/note="Vector: pMAK1; Site: Sfi IA; Site 2: Sfi IB; cDNA
was synthesized using SMART III oligo, CDS III oligo d(T
130 (Clontech) and Superscript II reverse transcriptase
(life Technology). cDNA clones were size fractionated and
directionally ligated in the Sfi IA restriction site at
5' end and Sfi IB at 3' end of pMAK1. pMAK1 was derived
from the plasmid pCDNA II (Invitrogen)."
BASE COUNT 343 a 149 c 208 g 101 t
ORIGIN
Query Match 13.6%; Score 36.8; DB 10; Length 801;
Best Local Similarity 54.4%; Pred. No. 24;
Matches 74; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 54 TCAATAAAATCATGATTCATCGAAGTTCATTCCTATACCAATATCTTTTAATAACA 113
DB 259 TTAATAAAATTAATGATTTTCCAAAAAATTTTATGGCTTTATATCCCCAAAAA 200
QY 114 TTTTATGCTCGGTTTATCGAATACCTTAACCTTCACCTTCGTAGCAGATCGTAGG 173
DB 199 TCATTACTTATGTTTATTTATTTTATTTTAAATGCAATCTCTTTCATTCGCAATCATC 140
QY 174 CATTCGCTATCAACTC 189

```

Sun Nov 16 12:57:03 2003

Db 139 CCTTGTGCTCTCCC 124

RESULT 15  
CC122282 809 bp DNA linear GSS 16-APR-2003  
LOCUS NDL.20H11.T7 Notre Dame Liverpool Aedes aegypti genomic clone  
DEFINITION NDL.20H11, genomic survey sequence.  
ACCESSION CC122282  
VERSION CC122282.1 GI:29991337  
KEYWORDS GSS.  
SOURCE Aedes aegypti (yellow fever mosquito)  
ORGANISM Aedes aegypti  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Aedes.  
REFERENCE 1 (bases 1 to 809)  
Loftus, B., Shetty, J., Knudson, D. and Severson, D.  
BAC end sequencing of Aedes aegypti  
Unpublished  
TITLE BAC end sequencing of Aedes aegypti  
JOURNAL  
COMMENT Other GSSs: NDL.20H11.SP6  
Contact: Brendan Loftus  
Department of Eukaryotic Genomics  
TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA  
Tel: 301-838-3543  
Fax: 301-838-0208  
Email: enta@tigr.org  
Library was provided by David Severson  
Seq primer: T7  
Class: BAC ends.  
Location/Qualifiers  
1. .809  
/organism="Aedes aegypti"  
/mol\_type="genomic DNA"  
/strain="Liverpool"  
/db\_xref="taxon:7159"  
/clone="NDL.20H11"  
/clone\_lib="Notre Dame Liverpool"  
/note="Vector: pCBAC1; Site 1: Hind III; The library was  
prepared from whole body tissue of newly hatched L1 larvae  
by David Severson at the University of Notre Dame and  
Hongbin Zhang"  
BASE COUNT 293 a 128 c 68 g 320 t  
ORIGIN

Query Match 13.6%; Score 36.8; DB 29; Length 809;  
Best Local Similarity 63.6%; Pred. No. 24;  
Matches 56; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 21 CTTGTTTTCTCCCTCTGTCGACGATCTCAATAAATCATTCATTCATCGA 80  
Db 450 CTCCTTTTCAATTTCTGTACTTCAATAATACATATAAAATATTGTTTCTCTA 509

QY 81 AGTTCATTCCTATACCATTCCTTAAT 108  
Db 510 TTTCGATTTTTTTTCAAAAATCTGTAAT 537

Search completed: November 15, 2003, 08:02:40  
Job time : 1806.19 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 15, 2003, 00:06:39 ; Search time 42.0547 Seconds  
(without alignments)  
2833.774 Million cell updates/sec

Title: US-09-928-457-88

Perfect score: 270

Sequence: 1 AATTCTCCGACGGGAGG.....CTCGAAACTCAATTCAATT 270

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA: \*  
1: /cgn2\_6/prodata/2/ina/5A\_COMB.seq: \*  
2: /cgn2\_6/prodata/2/ina/5B\_COMB.seq: \*  
3: /cgn2\_6/prodata/2/ina/6A\_COMB.seq: \*  
4: /cgn2\_6/prodata/2/ina/6B\_COMB.seq: \*  
5: /cgn2\_6/prodata/2/ina/PCTUS\_COMB.seq: \*  
6: /cgn2\_6/prodata/2/ina/backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	37.6	13.9	1230025	4	US-09-198-452A-1
C 2	36	13.3	4519	4	US-08-976-259-6
C 3	34.6	12.8	1437	4	US-09-004-838-37
C 4	34.6	12.8	5829	4	US-09-004-838-109
C 5	34.2	12.7	9566	4	US-09-328-352-3038
C 6	34	12.6	7218	1	US-08-232-463-14
C 7	33.2	12.3	14602	1	US-08-597-236-1
C 8	33.2	12.3	14602	1	US-08-746-682A-1
C 9	32.2	11.9	2781	3	US-08-749-522-4
C 10	31.2	11.6	1377	4	US-09-328-352-3423
C 11	31	11.5	2806	4	US-09-463-238-11
C 12	31	11.5	1830121	4	US-09-557-884-1
C 13	31	11.5	1830121	4	US-09-643-990A-1
C 14	30.8	11.4	1152	4	US-09-252-991A-12065
C 15	30.8	11.4	11580	4	US-09-334-220-4
C 16	30.8	11.4	24417	2	US-08-846-762-1
C 17	30.4	11.3	2577	2	US-08-209-521-25
C 18	30.4	11.3	2577	4	US-09-265-503B-135
C 19	30.4	11.3	5361	3	US-08-973-462-2
C 20	30.4	11.3	5152	3	US-08-973-462-1
C 21	30	11.1	501	2	US-08-928-692-56
C 22	30	11.1	501	4	US-09-339-972-56
C 23	30	11.1	1664976	4	US-08-916-421B-1
C 24	29.8	11.0	1465	4	US-09-004-838-100
C 25	29.8	11.0	1465	4	US-09-004-838-33
C 26	29.8	11.0	1474	4	US-09-004-838-36
C 27	29.8	11.0	4211	4	US-09-004-838-106

Sequence 1, Appli  
Sequence 39, Appli  
Sequence 27, Appli  
Sequence 3, Appli  
Sequence 3, Appli  
Sequence 5, Appli  
Sequence 5, Appli  
Sequence 5, Appli  
Sequence 3, Appli  
Sequence 3, Appli  
Sequence 21, Appli  
Sequence 87, Appli  
Sequence 6, Appli  
Sequence 5, Appli  
Sequence 5, Appli  
Sequence 8, Appli  
Sequence 28, Appli  
Sequence 137, App  
Sequence 459, App

28 29.8 11.0 1664976 4 US-08-916-421B-1  
C 29 29.4 10.9 1380 3 US-08-676-444-39  
C 30 29.4 10.9 1439 4 US-09-004-838-27  
C 31 29.4 10.9 3760 1 US-08-261-663A-3  
C 32 29.4 10.9 3760 5 PCT-US95-07754A-3  
C 33 29.4 10.9 3830 1 US-08-261-663A-5  
C 34 29.4 10.9 3830 5 PCT-US95-07754A-5  
C 35 29.4 10.9 7253 4 US-09-268-347-35  
C 36 29.4 10.9 7291 3 US-08-913-942-3  
C 37 29.4 10.9 10815 4 US-09-004-838-21  
C 38 29.4 10.9 13149 4 US-09-004-838-87  
C 39 29.2 10.8 966 1 US-08-257-073-6  
C 40 29.2 10.8 1814 1 US-08-097-829-5  
C 41 29.2 10.8 1814 1 US-08-577-403-5  
C 42 29.2 10.8 3056 4 US-09-708-200-8  
C 43 29.2 10.8 3065 2 US-08-209-521-28  
C 44 29.2 10.8 3065 4 US-09-265-503B-137  
C 45 29 10.7 592 3 US-09-385-982-459

#### ALIGNMENTS

#### RESULT 1

US-09-198-452A-1/c  
; Sequence 1, Application US/09198452A  
; Patent No. 6559294  
; GENERAL INFORMATION:  
; APPLICANT: Grifais, R  
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments thereof and uses thereof, in particular for the diagnosis, prevention of infection  
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention of infection  
; FILE REFERENCE: 9710-003-999  
; CURRENT APPLICATION NUMBER: US/09/198,452A  
; CURRENT FILING DATE: 1998-11-24  
; NUMBER OF SEQ ID NOS: 6849  
; SEQ ID NO 1

LENGTH: 1230025

TYPE: DNA

ORGANISM: Chlamydia pneumoniae

FEATURE:

NAME/KEY: misc feature

LOCATION: (1)-(15000)

OTHER INFORMATION: n-a or c or g or t

NAME/KEY: misc feature

LOCATION: (15001)..(30000)

OTHER INFORMATION: n-a or c or g or t

NAME/KEY: misc feature

LOCATION: (30001)..(45000)

OTHER INFORMATION: n-a or c or g or t

NAME/KEY: misc feature

LOCATION: (45001)..(60000)

OTHER INFORMATION: n-a or c or g or t

NAME/KEY: misc feature

LOCATION: (60001)..(75000)

OTHER INFORMATION: n-a or c or g or t

NAME/KEY: misc feature

LOCATION: (75001)..(90000)

OTHER INFORMATION: n-a or c or g or t

NAME/KEY: misc feature

LOCATION: (90001)..(105000)

OTHER INFORMATION: n-a or c or g or t

NAME/KEY: misc feature

LOCATION: (105001)..(120000)

OTHER INFORMATION: n-a or c or g or t

NAME/KEY: misc feature

LOCATION: (120001)..(135000)

OTHER INFORMATION: n-a or c or g or t

NAME/KEY: misc feature

LOCATION: (135001)..(150000)

OTHER INFORMATION: n-a or c or g or t

NAME/KEY: misc feature

LOCATION: (150001)..(165000)



LOCATION: (885001)...(900000)  
OTHER INFORMATION: nsa or c or g or t  
NAME/KEY: misc\_feature  
LOCATION: (900001)...(915000)  
OTHER INFORMATION: nsa or c or g or t  
NAME/KEY: misc\_feature

Query Match  
Best Local Similarity 13.9%; Score 37.6; DB 4; Length 1230025;  
Matches 76; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

70 GATTTCATCGAAGTTTCATTCTATACCATTTATCTTAATAACGATTTTATGCTCCGGTTT 129  
929463 GATGCTTCGGCTCTGAAGCCACACACCATTTATCTGTATATGATGGCCCTTGTCCCCC 929404

130 ATCGATTAACCTAATTCATCTCCGTAGACATGATCGTAGGATTCGGTATCAACTC 189  
929403 TCCCTAAAGTTTCGATTTCTATTTTCATCGCGCCAGCATCTAAGGAATTTCTCAATCAGTTC 929344

190 GGCAATCGCAGGAACAGTGT 209  
929343 TTTAACAACAGAACGGAGT 929324

RESULT 2  
US-08-976-259-6  
Sequence 6, Application US/08976259  
Patent No. 6316609  
GENERAL INFORMATION:  
APPLICANT: Dillon, Patrick J.  
APPLICANT: Choi, Gil H.  
APPLICANT: Welch, Rodney A.  
TITLE OF INVENTION: Nucleotide Sequence of Escherichia coli  
Patent No. 6316609  
NUMBER OF SEQUENCES: 142  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
STREET: 1100 New York Ave, N.W., Suite 600  
City: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20005-3934  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/976,259  
FILING DATE: Herewith  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/031,626 AND US 60/061,953  
ATTORNEY/AGENT INFORMATION:  
NAME: Steffe, Eric K.  
REGISTRATION NUMBER: 36,688  
REFERENCE/DOCKET NUMBER: 1488.0740002/EKS/CBM  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 371-2600  
TELEFAX: (202) 371-2540  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4519 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-976-259-6

Query Match  
Best Local Similarity 13.38; Score 36; DB 4; Length 4519;  
Matches 88; Conservative 0; Mismatches 65; Indels 3; Gaps 1;

47 CCGATTCTCAATAAATAAATCAATTCATTTCATCGAAGTTTCATTCCTATACCATTTATCTTTA 106

95 CCAATAGTCAACCAACCTCTCTTCAAATTCATCGGTTGTATACCTAATCCATCATCT--- 151  
107 ATACGATTTTATGCTCCGGTTTATCGAATAAATCACTTCCACTTCGTTAGCACATGCA 166  
152 CTCAGATAGAAGATTTTCTTCTTAAATAAATCACTTCGACATATATCAGCATAGGCA 211  
167 TCGTAGGATTCGCTATCACTCGGCAATCGCAGGA 202  
212 TCATGAGCATTTTAAATAAATCACTCAAGGAGTA 247

RESULT 3  
US-09-004-838-37/c  
Sequence 37, Application US/09004838  
Patent No. 6350933  
GENERAL INFORMATION:  
APPLICANT: Michelmore, Richard W.  
APPLICANT: Shen, Kathy  
APPLICANT: Meyers, Blake  
TITLE OF INVENTION: Procedures and Materials for  
TITLE OF INVENTION: Confering Pest Resistance in Plants  
NUMBER OF SEQUENCES: 140  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
City: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/004,838  
FILING DATE: 09-JAN-1998  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/781,734  
FILING DATE: 10-JAN-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Einhorn, Gregory P.  
REGISTRATION NUMBER: 38,440  
REFERENCE/DOCKET NUMBER: 023070-078810US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 37:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1437 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
FEATURE:  
NAME/KEY: -  
LOCATION: 1, 1437  
OTHER INFORMATION: /note= "RLG2K"  
US-09-004-838-37

Query Match  
Best Local Similarity 12.8%; Score 34.6; DB 4; Length 1437;  
Matches 61; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

44 CGACGATTCCTCAATAAATAAATCAATTCATTCGATTCGAAAGTTTCATTCGATTCCTATACCATTTATCT 103  
1422 CCACCAACTCTCATATAAAGCTTTCAAGTTTCCAAATTTCTTTAGGACACCATTTATCT 1363  
104 TTAATAACGATTTTATGCTCCGGTTTATCGAATAACCTTAATCTTC 148  
1362 ATACGAGACCATCACAATTTGTCAAATCTAGTACCTTACTCTTC 1318

```

RESULT 4
US-09-004-838-109/c
; Sequence 109, Application US/09004838
; Patent No. 6350933
; GENERAL INFORMATION:
; APPLICANT: Michelmore, Richard W.
; APPLICANT: Shen, Kathy
; APPLICANT: Meyers, Blake
; TITLE OF INVENTION: Procedures and Materials for
; TITLE OF INVENTION: Conferring Pest Resistance in Plants
; NUMBER OF SEQUENCES: 140
; CORRESPONDENCE ADDRESS:
; ADDRESS: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/09/004,838
; FILING DATE: 09-JAN-1998
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/781,734
; FILING DATE: 10-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Einhorn, Gregory P.
; REGISTRATION NUMBER: 38,440
; REFERENCE/DOCKET NUMBER: 023070-078810US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 109:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5829 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..5829
; OTHER INFORMATION: /note= "RG2K"
US-09-004-838-109

Query Match 12.8%; Score 34.6; DB 4; Length 5829;
Best Local Similarity 58.1%; Pred. No. 0.36;
Matches 61; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 44 CGACCGATTCTCAATAAAATCATTTGATTCATCGAAGTTTCATTCCTATACCATTTATCT 103
DB 2678 CCACCAATCTCATATAAGCTTCITCAAGTTTCACCAATTTCTTTAGGACACCATTTATCT 2619

QY 104 TTAATAAGATTTTATGCTCCGGTTTATCGAATAACCTTAATCTTC 148
DB 2618 ATACGAAGACCATCAACAATTTGTCAAAATCTAGTACCCCTTAGCTCC 2574

RESULT 5
US-09-328-352-3038/c
; Sequence 3038, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS

```

```

; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 3038
; LENGTH: 966
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-3038

Query Match 12.7%; Score 34.2; DB 4; Length 966;
Best Local Similarity 49.8%; Pred. No. 0.27;
Matches 113; Conservative 0; Mismatches 113; Indels 1; Gaps 1;

QY 28 TTCCTCCCTTCTGTTCCGACCGATTCTCAATAAAATCATTTGATTCATTCGAAGTTTCAT 87
DB 885 TTCCTCGCTCTTGTGTGATAGCAACAATACTTTAGAAATCTTTCAATCCAGCAAGTCTG 826
QY 88 TCCTATACCATTAICTTTTAATACGATTTTATGCTCCGGTTTATCGAATAACCTTAATTC 147
DB 825 GATCCGACCAAGAAATACCTACAGCAATGTAAGCTCAGGTGCAACGATTTTAC-CAGTTT 767
QY 148 CACTTCCGTAGCACATGATCGTAGGATTCGCTATCAACTCGGCAATCGCAGGAACAGT 207
DB 766 GACCAACTTGAAGTCGTTAGGTACAAAGCCTGCATCAACTGCTGCAGGTGAACACCTT 707
QY 208 GTGCGAATACAAATCTTTACACCCAAATGTTTCGATTACGGTTGGTCTG 254
DB 706 GTGCTGCACCAAGCTTGTCACTATGCTCAAGTACTTCTGTTGGTAG 660

RESULT 6
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22133-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs

```





DB 10630 GCAACATGCTTCTTACCATAGAAAAACCTCCCAATGTACATTT 10585

RESULT 8

US-08-746-682A-1/c  
 ; Sequence 1, Application US/08746682A  
 ; Patent No. 5786184  
 ; GENERAL INFORMATION:  
 ; APPLICANT: STINGELE, Francesca  
 ; APPLICANT: MOLLET, Beat  
 ; TITLE OF INVENTION: LACTIC BACTERIA PRODUCING  
 ; TITLE OF INVENTION: EXOPOLYSACCHARIDES  
 ; NUMBER OF SEQUENCES: 19  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Pennie & Edmonds  
 ; STREET: 1155 Avenue of the Americans  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: U.S.A.  
 ; ZIP: 10036  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent in Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/746.682A  
 ; FILING DATE: 14-NOV-1996  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/597,236  
 ; FILING DATE: 20-JUN-1995  
 ; APPLICATION NUMBER: EP 95201669.9  
 ; FILING DATE: 20-JUN-1995  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Fanucci A., Allan  
 ; REGISTRATION NUMBER: 30256  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 14602 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 352..1803  
 ; OTHER INFORMATION: /product= "epsa"  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 1807..2535  
 ; OTHER INFORMATION: /product= "epsb"  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 2547..3239  
 ; OTHER INFORMATION: /product= "epsc"  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 3249..3995  
 ; OTHER INFORMATION: /product= "epsd"  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 4051..4731  
 ; OTHER INFORMATION: /product= "epse"  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 4898..5854  
 ; OTHER INFORMATION: /product= "epsf"  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 6425..7540  
 ; OTHER INFORMATION: /product= "epsg"

FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 7736..8212  
 OTHER INFORMATION: /product= "epsh"  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 8221..9192  
 OTHER INFORMATION: /product= "epsi"  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 9285..10364  
 OTHER INFORMATION: /product= "epsj"  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 10392..11339  
 OTHER INFORMATION: /product= "epsk"  
 FEATURE:  
 NAME/KEY: misc.feature  
 LOCATION: 11302..12222  
 OTHER INFORMATION: /product= "CDS (eps L) covering CDS  
 OTHER INFORMATION: (eps k) on nucleotides 10392-11339"  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 12233..13651  
 OTHER INFORMATION: /product= "epsm"  
 FEATURE:  
 NAME/KEY: misc.feature  
 LOCATION: 13732..14305  
 OTHER INFORMATION: /function= "CDS on the  
 OTHER INFORMATION: complementary strand"  
 OTHER INFORMATION: /product= "off2"  
 FEATURE:  
 NAME/KEY: terminator  
 LOCATION: 230..252  
 FEATURE:  
 NAME/KEY: promoter  
 LOCATION: 274..302  
 FEATURE:  
 NAME/KEY: RBS  
 LOCATION: 340..345  
 US-08-746-682A-1  
 Query Match 12.3%; Score 33.2; DB 1; Length 14602;  
 Best Local Similarity 50.0%; Pred No. 1.3;  
 Matches 83; Conservative 0; Mismatches 83; Indels 0; Gaps 0;  
 QY 59 TAAAAATCATGATTTCATCGAAGTTTCATTCCTATACCATTCCTTTTAAACGATTTTA 118  
 Db 10750 TTAGGATAACTATTTTCAAAAAAATAAATACTATTTCATTATCTTTGATTGCGATATCA 10691  
 QY 119 TGCTCCGGTTTATCGAATACCTACCTCCATTCCTCGTAGCACATCGTAGGCATTC 178  
 Db 10690 TCTCCAGATAAAAAATGAAAAATAGAAATATTCCTGTATACGTCGACATTCCTAATAGC 10631  
 QY 179 GCTATCAACTCGCAATCGCAGGAACAGTGTGCGAATACAAATCTTT 224  
 Db 10630 GCAACATTCCTTCTACCATAGAAAAACCTCCCAATGTACATTT 10585  
 RESULT 9  
 US-08-749-522-4  
 ; Sequence 4, Application US/08749522  
 ; Patent No. 6096950  
 ; GENERAL INFORMATION:  
 ; APPLICANT: John, Maliyakal  
 ; TITLE OF INVENTION: FIBER-SPECIFIC PROMOTERS  
 ; NUMBER OF SEQUENCES: 8  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Quarles & Brady  
 ; STREET: 411 East Wisconsin Avenue  
 ; CITY: Milwaukee  
 ; STATE: WI  
 ; COUNTRY: U.S.A.

```

QY      134 AATAACCTAACT 145
      |||||
Db      1148 TATCATAACT 1159

RESULT 11
US-09-463-238-11/c
; Sequence 11, Application US/09463238
; Patent No. 6469230
; GENERAL INFORMATION:
; APPLICANT: Edwards, Elizabeth A
; APPLICANT: Smith, Alison M
; APPLICANT: Bustos Guillen, Regla
; APPLICANT: Martin, Catherine R
; APPLICANT: Plant Bioscience Limited
; TITLE OF INVENTION: Starch Debranching Enzymes
; FILE REFERENCE: 97.118
; CURRENT APPLICATION NUMBER: US/09/463,238
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: PCT/GB98/02280
; PRIOR FILING DATE: 1998-07-30
; PRIOR APPLICATION NUMBER: GB 9716185.5
; PRIOR FILING DATE: 1997-07-31
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 2806
; TYPE: DNA
; ORGANISM: Solanum tuberosum
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (822, 826, 2707 and 2797)
; OTHER INFORMATION: n = a or g or c or t
US-09-463-238-11

Query Match      11.5%; Score 31; DB 4; Length 2806;
Best Local Similarity 48.8%; Pred. No. 3.6;
Matches 82; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

QY      1 AATCTTCGCGAGGGAGCGTTGTTTTCTCTCCCTCTCTGTCGACCGCATTCFCAAATA 60
      |||||
Db      2796 AATCGAGATCGGTACCATCGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 2737

QY      61 AAAATCATTGATTCATCGAAGTCATTCCTATACCATTATCTTTAATAACGATTTTATG 120
      |||||
Db      2736 TTTTTCGCGACAGACAATTCACTANTATTTTATATATAAATCAATTAATAA 2677

QY      121 CTCGGTTTATCGAATAACTACTTCCTCCACTTCGGTAGCAGTCATC 168
      |||||
Db      2676 ACAAGGTGATCTTATAATCATCAGCTTCTTCCTCACAATTTTATC 2629

RESULT 12
US-09-557-884-1/c
; Sequence 1, Application US/09557884
; Patent No. 6506581
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: The Nucleotide sequence of
; the Haemophilus influenzae Rd Genome, Fragments
; Thereof, and Uses Thereof
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: MD
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3 1/2 inch diskette
; COMPUTER: Dell Pentium
; OPERATING SYSTEM: MS DOS v6.22

```

```
;
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/557,884
; FILING DATE: 25-Apr-2000
; CLASSIFICATION: <unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/476,102
; FILING DATE: JUN-5-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB186P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-309-8504
; TELEFAX: 301-309-8439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1830121 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1

Query Match 11.5%; Score 31; DB 4; Length 1830121;
Best Local Similarity 56.3%; Pred. No. 26;
Matches 58; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 28 TTCTCCCTCTGTTCCGACGATTCTCAATAAAATCAATGATTTTCATCGAAGTTTCAT 87
Db 809911 TACATCCCTAGTACTATGACAGCTAATAATCACCATTAAGATACCAACATTTTTCAT 809852

QY 88 TCCTATACCATATCTTTAATAACGATTTTATGCTCCGGTTTA 130
Db 809851 TGATAAACCATAACTTTAAGTAAGAAGTCAGCCTGTGGTTTA 809809

RESULT 13
US-09-643-990A-1/c
; Sequence 1, Application US/09643990A
; Patent No. 6528289
; GENERAL INFORMATION:
; APPLICANT: Robert D. Fleischmann
; Mark D. Adams
; Owen White
; Hamilton O. Smith
; J. Craig Venter
; TITLE OF INVENTION: The Nucleotide sequence of
; the Haemophilus influenzae Rd Genome, Fragments
; Thereof, and Uses Thereof
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville,
; STATE: MD
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3 1/2 inch diskette
; COMPUTER: Dell Pentium
; OPERATING SYSTEM: MS DOS v6.22
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/643,990A
; FILING DATE: 23-AUG-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/487,429
; FILING DATE: 1995-06-07
; APPLICATION NUMBER: 08/426,787
; FILING DATE: 1995-04-21
; ATTORNEY/AGENT INFORMATION:
```

```
;
; NAME: Kenley K. Hoover
; REGISTRATION NUMBER: 40,302
; REFERENCE/DOCKET NUMBER: PB186P1C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-610-5790
; TELEFAX: 310-309-8439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1830121 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1

Query Match 11.5%; Score 31; DB 4; Length 1830121;
Best Local Similarity 56.3%; Pred. No. 26;
Matches 58; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 28 TTCTCCCTCTGTTCCGACGATTCTCAATAAAATCAATGATTTTCATCGAAGTTTCAT 87
Db 809911 TACATCCCTAGTACTATGACAGCTAATAATCACCATTAAGATACCAACATTTTTCAT 809852

QY 88 TCCTATACCATATCTTTAATAACGATTTTATGCTCCGGTTTA 130
Db 809851 TGATAAACCATAACTTTAAGTAAGAAGTCAGCCTGTGGTTTA 809809

RESULT 14
US-09-252-991A-12085/c
; Sequence 12085, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 12085
; LENGTH: 1152
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-12085

Query Match 11.4%; Score 30.8; DB 4; Length 1152;
Best Local Similarity 50.0%; Pred. No. 3.1;
Matches 77; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 55 CAATATAAAATCATTCATTCATCGAAGTTTCATTCCTATACCAATATCTTTAATAACGAT 114
Db 423 CAATCGAAGATCACGAATTTCTCCAGTTTATTACATGATGTAATAACAAACCCCA 364

QY 115 TTTATGCTCCGGTTTATCGAATAACCTAACCTCCACTTCCTAGCACATGATCGTAGGC 174
Db 363 ATATTTTCAATCTTTTCAATATTACTGCTGCAAGTTTCGCTATTCCAGCGCGGTCCAC 304

QY 175 ATTCGCTATCAACTCGGCAATCGGAGGAACAGTG 208
Db 303 ATGCACACCAAGTGGGCGCAAGCCCAAGATCCTTG 270

RESULT 15
US-09-334-220-4/c
; Sequence 4, Application US/09334220
; Patent No. 6323177
; GENERAL INFORMATION:
; APPLICANT: St. Jude's Children's Research Hospital
```

; APPLICANT: Curran, Thomas  
; APPLICANT: D'Arcangelo, Gabriella  
; TITLE OF INVENTION: INTERACTION OF REELIN WITH VERY LOW  
; TITLE OF INVENTION: DENSITY LIPOPROTEIN (VLDL) RECEPTOR FOR SCREENING AND  
; TITLE OF INVENTION: THERAPIES  
; FILE REFERENCE: 2427/0F704  
; CURRENT APPLICATION NUMBER: US/09/334,220  
; CURRENT FILING DATE: 1999-06-16  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 4  
; LENGTH: 11580  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-09-334-220-4

Query Match 11.4%; Score 30.8; DB 4; Length 11580;  
Best Local Similarity 50.0%; Pred. No. 6.4;  
Matches 77; Conservative 0; Mismatches 77; Indels 0; Gaps 0;  
QY 38 CTGTTCCGACCGATTCTCAATAAAATCATTGATTTCATCGAAGTTGATTCCTATACCA 97  
Db 3933 CTCATCTGATTCTAGGGTAATCAAAAGCTGGCTTCTATAAAAGTTCTGAGGTAAAGTT 3874  
QY 98 TTATCTTTAATAACGATTTTATGCTCCGGTTTATCGAATAACCTAACTTCCACTTCCGTA 157  
Db 3873 GGATTGATAACTGGGATGATCTGCTTCTGCTTCTCGGACAGAAATGATGATGTCATCGACT 3814  
QY 158 GCACATGCATCGTAGGCATTGCTCAACTCGG 191  
Db 3813 GCCCACTGGTCATAGTCTCCCTGAGAACGG 3780

Search completed: November 15, 2003, 08:09:27  
Job time : 51.0547 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 15, 2003, 00:09:34 ; Search time 152.896 Seconds  
(without alignments)  
5771.357 Million cell updates/sec

Title: US-09-928-457-88

Perfect score: 270

Sequence: 1 AATTCTCGCAGCGGGAGG.....CTCGAATCAATTCATT 270

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2169961 seqs, 163402185 residues

Total number of hits satisfying chosen parameters: 4339922

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:\*  
1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW PUB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW PUB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW PUB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW PUB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:\*  
13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:\*  
14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq:\*  
16: /cgn2\_6/ptodata/1/pubpna/US60\_NEW PUB.seq:\*  
17: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	270	100.0	270	10	US-09-928-457-88
2	121.8	45.1	269	10	US-09-928-457-77
3	36	13.3	4519	9	US-09-956-004-6
4	34.6	12.8	10151	12	US-10-311-455-2406
5	34.4	12.7	543	9	US-09-864-761-7591
6	34	12.6	15767	12	US-10-311-455-1179
7	34	12.6	15767	12	US-10-240-485-105
8	33.4	12.4	285	11	US-09-991-936-1827
9	33.4	12.4	285	11	US-09-991-936-1853
10	33.2	12.3	3673778	12	US-10-312-841-1
11	33	12.2	5360	12	US-10-204-708-66
12	33	12.2	5360	12	US-10-311-455-1910
13	32.8	12.1	7811	12	US-10-311-455-1255
14	32.6	12.1	2000	10	US-09-938-842A-4571
15	32.6	12.1	5218	12	US-10-311-455-1239
16	32.6	12.1	6114	12	US-10-032-595-6119

17	32.4	12.0	6118	12	US-10-311-455-1004	Sequence 1004, Ap
C 18	32.4	12.0	6341	12	US-10-311-455-1617	Sequence 1617, Ap
C 19	32.2	11.9	6608	12	US-10-311-455-2007	Sequence 2007, Ap
C 20	32.2	11.9	16173	12	US-10-240-485-22	Sequence 22, Appl
C 21	32	11.9	8446	12	US-10-311-455-1643	Sequence 1643, Ap
C 22	32	11.9	85548	14	US-10-175-523-75	Sequence 75, Appl
C 23	31.6	11.7	11125	8	US-08-781-986A-171	Sequence 171, Appl
C 24	31.6	11.7	12393	12	US-10-311-455-1235	Sequence 1235, Ap
C 25	31.4	11.6	423	12	US-10-027-632-77344	Sequence 77344, A
C 26	31.4	11.6	423	12	US-10-027-632-300417	Sequence 300417, A
C 27	31.4	11.6	423	13	US-10-027-632-77344	Sequence 77344, A
C 28	31.4	11.6	423	13	US-10-027-632-300417	Sequence 300417, A
C 29	31.4	11.5	8413	9	US-09-801-574-7	Sequence 7, Appli
C 30	31	11.5	155	10	US-09-867-701-4815	Sequence 4815, Ap
C 31	31	11.5	1396	10	US-09-938-842A-3999	Sequence 3999, Ap
C 32	31	11.5	1984	10	US-09-887-576-38	Sequence 38, Appl
C 33	31	11.5	2000	10	US-09-938-842A-2883	Sequence 2883, Ap
C 34	31	11.5	3434	12	US-10-148-351-45	Sequence 45, Appl
C 35	31	11.5	4049	12	US-10-148-351-66	Sequence 66, Appl
C 36	31	11.5	6930	10	US-09-070-927A-18	Sequence 18, Appl
C 37	31	11.5	1930121	14	US-10-329-960-1	Sequence 1, Appli
C 38	31	11.5	2140405	12	US-10-027-632-76212	Sequence 76212, A
C 39	31	11.5	2140405	13	US-10-027-632-76212	Sequence 76212, A
C 40	31	11.5	3186778	12	US-10-027-632-174961	Sequence 174961, A
C 41	31	11.5	3186778	13	US-10-027-632-174961	Sequence 174961, A
C 42	30.8	11.4	1188	12	US-10-027-632-265139	Sequence 265139, A
C 43	30.8	11.4	1188	12	US-10-027-632-265140	Sequence 265140, A
C 44	30.8	11.4	1188	13	US-10-027-632-265139	Sequence 265139, A
C 45	30.8	11.4	1188	13	US-10-027-632-265140	Sequence 265140, A

ALIGNMENTS

RESULT 1

US-09-928-457-88  
; Sequence 88, Application US/09928457  
; Patent No. US20020164603A1

GENERAL INFORMATION:

APPLICANT: DNA, specific proteins and peptides  
TITLE OF INVENTION: Of the Neisseria meningitidis species bacteria, method  
TITLE OF INVENTION: for obtaining them and their biological application.  
NUMBER OF SEQUENCES: 99  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30 (OEB)

SEQUENCE CHARACTERISTICS:

SEQUENCE FOR SEQ ID NO: 88:  
LENGTH: 270 base pairs  
TYPE: nucleotide  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO

SEQUENCE CHARACTERISTICS:

SEQUENCE FOR SEQ ID NO: 88:  
LENGTH: 270 base pairs  
TYPE: nucleotide  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO

SEQUENCE CHARACTERISTICS:

SEQUENCE FOR SEQ ID NO: 88:  
LENGTH: 270 base pairs  
TYPE: nucleotide  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO

SEQUENCE CHARACTERISTICS:

SEQUENCE FOR SEQ ID NO: 88:  
LENGTH: 270 base pairs  
TYPE: nucleotide  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO

SEQUENCE CHARACTERISTICS:

SEQUENCE FOR SEQ ID NO: 88:  
LENGTH: 270 base pairs  
TYPE: nucleotide  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO

SEQUENCE CHARACTERISTICS:

SEQUENCE FOR SEQ ID NO: 88:  
LENGTH: 270 base pairs  
TYPE: nucleotide  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO

SEQUENCE CHARACTERISTICS:

SEQUENCE FOR SEQ ID NO: 88:  
LENGTH: 270 base pairs  
TYPE: nucleotide  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO

SEQUENCE CHARACTERISTICS:

SEQUENCE FOR SEQ ID NO: 88:  
LENGTH: 270 base pairs  
TYPE: nucleotide  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO

SEQUENCE CHARACTERISTICS:

SEQUENCE FOR SEQ ID NO: 88:  
LENGTH: 270 base pairs  
TYPE: nucleotide  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO

SEQUENCE CHARACTERISTICS:

SEQUENCE FOR SEQ ID NO: 88:  
LENGTH: 270 base pairs  
TYPE: nucleotide  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO

SEQUENCE CHARACTERISTICS:

SEQUENCE FOR SEQ ID NO: 88:  
LENGTH: 270 base pairs  
TYPE: nucleotide  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO

QY	61	AAAATCATTTGATTTCATCGAAGTTTCATTCCTATACCATTATCTTTAATAACGATTTTATG	120
Db	61	AAAATCATTTGATTTCATCGAAGTTTCATTCCTATACCATTATCTTTAATAACGATTTTATG	120
QY	121	CTCCGGTTTATCGAATAACCTAACTTCACATTCGGTAGCACATCGCATCGTAGGCATTCGC	180
Db	121	CTCCGGTTTATCGAATAACCTAACTTCACATTCGGTAGCACATCGCATCGTAGGCATTCGC	180
QY	181	TATCAACTCGGCAATCGCAGGAACAGTGTGCGAATACAATCTTTACACCCAAATGTTCGA	240
Db	181	TATCAACTCGGCAATCGCAGGAACAGTGTGCGAATACAATCTTTACACCCAAATGTTCGA	240
QY	241	TTACGGTTGGCTCGAAACTCAATTTCAAAT	270
Db	241	TTACGGTTGGCTCGAAACTCAATTTCAAAT	270

## RESULT 2

```

US-09-928-457-77/c
; Sequence 77, Application US/09928457
; Patent No. US20020164603A1
; GENERAL INFORMATION:
;
; APPLICANT:
;
; TITLE OF INVENTION: DNA, specific proteins and peptides
;
; TITLE OF INVENTION: of the Neisseria meningitidis species bacteria, method
;
; TITLE OF INVENTION: for obtaining them and their biological application.
;
; NUMBER OF SEQUENCES: 99
;
; COMPUTER READABLE FORM:
;
; MEDIUM TYPE: Floppy disk
;
; COMPUTER: IBM PC compatible
;
; OPERATING SYSTEM: PC-DOS/MS-DOS
;
; SOFTWARE: Patent In Release #1.0, Version #1.30 (OEB)
;
; CURRENT APPLICATION DATA:
;
; APPLICATION NUMBER: US/09/928,457
;
; FILING DATE: 2001-08-14
;
; PRIOR APPLICATION DATA:
;
; APPLICATION NUMBER: 09/214,759
;
; FILING DATE: 199-12-10
;
; INFORMATION FOR SEQ ID NO: 77:
;
; SEQUENCE CHARACTERISTICS:
;
; LENGTH: 269 base pairs
;
; TYPE: nucleotide
;
; STRANDEDNESS: single
;
; TOPOLOGY: linear
;
; MOLECULE TYPE: DNA (genomic)
;
; HYPOTHETICAL: NO
;
; ANTI-SENSE: NO
;
; US-09-928-457-77

```

### RESULT 3

US-09-956-004-6  
; Sequence 6, Application US/09956004  
; Patent No. US20020072595A1  
; GENERAL INFORMATION:  
; APPLICANT: Patrick J. Dillon et al

```

; TITLE OF INVENTION: Nucleotide Sequences of Escherichia coli Pathogenicity Islands
; FILE REFERENCE: PB9324D1
; CURRENT APPLICATION NUMBER: US/09/956,004
; CURRENT FILING DATE: 2001-09-20
; PRIOR APPLICATION NUMBER: 08/976,259
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: 60/061,953
; PRIOR FILING DATE: 1997-10-14
; PRIOR APPLICATION NUMBER: 60/031,626
; PRIOR FILING DATE: 1996-11-22
; NUMBER OF SEQ ID NOS: 142
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 4519
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (3483)..(3483)
; OTHER INFORMATION: n equals a, t, g, or c
; NAME/KEY: misc.feature
; LOCATION: (3487)..(3487)
; OTHER INFORMATION: n equals a, t, g, or c
; NAME/KEY: misc.feature
; LOCATION: (4292)..(4292)
; OTHER INFORMATION: n equals a, t, g, or c
; NAME/KEY: misc.feature
; LOCATION: (4318)..(4318)
; OTHER INFORMATION: n equals a, t, g, or c
; NAME/KEY: misc.feature
; LOCATION: (4329)..(4329)
; OTHER INFORMATION: n equals a, t, g, or c
; US-09-956-004-6

Query Match          13.3%; Score 36; DB 9; Length 4519;
Best Local Similarity 56.4%; Pred. No. 10;
Matches 88; Conservative 0; Mismatches 65; Indels 3; Gaps 1;

QY      47  CGGATTCGAAATAAAATCATGTTGTCATCGGAAGTTCATCTCTATACCAATTCCTTTA 106
Db      95  CCAATAGTCAACCACTCTCTTTCAAATTCATCGGTTGTGCATACCTAATCCATCATCT 151
QY      107 ATAAAGCATTTTATGCTCCGGTTATTCGAATAACCTAACTTCACCTCCGTAGCACATGCA 166
Db      152 CTCACAGTAAGAGATTTTCTTCTCTAAAAAATCAACTTCGACATTCAGCATAGCA 211
QY      167 TCGTAGGCAATTCGCTATCATCACTCGGCAATCCGAGGA 202
Db      212 TCATGAGCATTTTAAATAAATCACTCAAGGCAGTA 247

```



```
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 9495..9552
; OTHER INFORMATION: n is a or g or c or t
US-10-311-455-2406

Query Match
Best Local Similarity 12.8%; Score 34.6; DB 12; Length 10151;
Matches 61; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

Qy 41 TTCGACCGATTCTCAATAAAATCATGATTCATCGAAGTTCAATTCCTATACCATTA 100
Db 1152 TCCCAAAACCATCTCTCTCACTCAACCCCTTCATAAATAAATACTACTCTCAAAACCATCA 1093

Qy 101 TCTTTAATAACGATTTTATGCTCCGGTTTATCGAATAACCTAACT 145
Db 1092 TACCTAACAAATTTTATTTATTTTATTTTCTTAATAACAAATTT 1048

RESULT 5
US-09-864-761-7691/c
; Sequence 7691, Application US/09864761
; Patent No. US20030048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aemica-X-1
; CURRENT APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/006656
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006657
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006654
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006659
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006655
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006658
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006653
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006652
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006651
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,697
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
```

```
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annotmax Sequence Listing Engine vers. 1.1
; SEQ ID NO 7691
; LENGTH: 543
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC016648.3
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.84
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.9
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.81
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.89
US-09-864-761-7691

Query Match
Best Local Similarity 12.7%; Score 34.4; DB 9; Length 543;
Matches 74; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

Qy 3 TTCTTCGACGGGGAGGCTTGTTTTTCTTCCCTTCTGTTCCGACCGATTCTCAATAAA 62
Db 508 TTAATCGACAAATGTTTGTATAGTTTTCACAAATTCCTCTGTTATTGATTTCCAGTTTA 449

Qy 63 AATCATGATTCATCGAAGTTTCATTCCTATACCATATCTTTTAATACGATTTTATGCT 122
Db 448 TTCCATGTGTCAGAGAAGATGCTTGATATTAATCTTCAAAATTTTGAATGTTTAAAGAT 389

Qy 123 CCGGTTTATCGAATAACCTA 142
Db 388 TGGTTTTTCACATAACATA 369

RESULT 6
US-10-311-455-1179/c
; Sequence 1179, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1179
; LENGTH: 15767
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1179

Query Match
Best Local Similarity 12.6%; Score 34; DB 12; Length 15767;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

Qy 41 TTCGACCGATTCTCAATAAAATCATGATTCATCGAAGTTTCATTCCTATACCATTA 100
Db 2231 TTCCAAATTTACTTTTATAACAAAATTTTAACTTAAACCAATAATATAATTAACACTTA 2172

Qy 101 TCTTTAATAACGATTTTATGCTCCGGTTTATCGAATAACCTAACTTCCACTTCCGTAGCA 160
Db 2171 ATTITTATTAATAATCTACTCTTTCTTTCCAAATAATCTCACAATCTCTTAACAACA 2112
```

QY 161 CA 162  
Db 2111 AA 2110

## RESULT 7

US-10-240-485-105/c  
; Sequence 105, Application US/10240485  
; Publication No. US20030149327A1  
; GENERAL INFORMATION:  
; APPLICANT: PISPENBROCK, Christian  
; APPLICANT: OLEK, Alexander  
; APPLICANT: BERLIN, Kurt  
; TITLE OF INVENTION: Diagnosis of Diseases Associated with  
; TITLE OF INVENTION: Metastasis  
; FILE REFERENCE: 5013.1007  
; CURRENT APPLICATION NUMBER: US/10/240,485  
; PRIOR FILING DATE: 2002-10-02  
; PRIOR APPLICATION NUMBER: PCI/EP01/03970  
; PRIOR FILING DATE: 2001-04-06  
; PRIOR APPLICATION NUMBER: DE 10019058.8  
; PRIOR FILING DATE: 2000-04-06  
; PRIOR APPLICATION NUMBER: DE 10019173.8  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: DE 10032529.7  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: DE 10043826.1  
; PRIOR FILING DATE: 2000-09-01  
; NUMBER OF SEQ ID NOS: 202  
; SEQ ID NO 105  
; LENGTH: 15767  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-240-485-105

Query Match 12.6%; Score 34; DB 12; Length 15767;

Best Local Similarity 54.9%; Pred. No. 61;  
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 41 TTCGACCGATTCCTCAATAAATAATCAATTTTCATCGAAGTTTCATCTCTATACCATTA 100  
Db 2231 TTCGAATTTACTTTTATTAACAAAATTTTAACTTAAACCAATAATATTAAATTACACTTA 2172

QY 101 TCTTTAAATACGATTTTATGCTCCGTTTATCGAATACCTTAACCTCCACTTCGTTAGCA 160  
Db 2171 ATTITTATTAATATCTACTCTTTCTTTTCCAAATAATCTCAATCTTCTTAAACAACA 2112

QY 161 CA 162  
Db 2111 AA 2110

## RESULT 8

US-09-991-936-1627  
; Sequence 1627, Application US/09991936  
; Publication No. US20030073827A1  
; GENERAL INFORMATION:  
; APPLICANT: Brandt, Kevin S.  
; APPLICANT: Gaines, Patrick J.  
; APPLICANT: Stinchcomb, Dan T.  
; APPLICANT: Wisniewski, Nancy  
; TITLE OF INVENTION: FLEA HEAD, NERVE CORD, HINDGUT AND MALPIGHIAN TUBULE  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES, PROTEINS AND USES THEREOF  
; FILE REFERENCE: FC-6-C1  
; CURRENT APPLICATION NUMBER: US/09/991,936  
; CURRENT FILING DATE: 2001-11-21  
; PRIOR APPLICATION NUMBER: US/09/543,668  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: 60/128,704  
; PRIOR FILING DATE: 1999-04-09

; NUMBER OF SEQ ID NOS: 1959  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1627  
; LENGTH: 285  
; TYPE: DNA  
; ORGANISM: Ctenocephalides felis  
US-09-991-936-1627

Query Match 12.4%; Score 33.4; DB 11; Length 285;

Best Local Similarity 62.7%; Pred. No. 17;  
Matches 52; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 61 AAAATCATTGATTTTCATCGAAGTTTCATCTCTATACCATTTCTTTAAACGATTTTATG 120  
Db 67 AAAATCATTGATTTATTTATAAATCATAGCAAAATTTATCTCGATATTTATTATAGC 126

QY 121 CTCGGTTTATCGAATAACCTAA 143  
Db 127 GTCATATTGATCTAATAAAGAA 149

## RESULT 9

US-09-991-936-1853  
; Sequence 1853, Application US/09991936  
; Publication No. US20030073827A1  
; GENERAL INFORMATION:  
; APPLICANT: Brandt, Kevin S.  
; APPLICANT: Gaines, Patrick J.  
; APPLICANT: Stinchcomb, Dan T.  
; APPLICANT: Wisniewski, Nancy  
; TITLE OF INVENTION: FLEA HEAD, NERVE CORD, HINDGUT AND MALPIGHIAN TUBULE  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES, PROTEINS AND USES THEREOF  
; FILE REFERENCE: FC-6-C1  
; CURRENT APPLICATION NUMBER: US/09/991,936  
; CURRENT FILING DATE: 2001-11-21  
; PRIOR APPLICATION NUMBER: US/09/543,668  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: 60/128,704  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 1959  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1853  
; LENGTH: 285  
; TYPE: DNA  
; ORGANISM: Ctenocephalides felis  
US-09-991-936-1853

Query Match 12.4%; Score 33.4; DB 11; Length 285;

Best Local Similarity 62.7%; Pred. No. 17;  
Matches 52; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 61 AAAATCATTGATTTTCATCGAAGTTTCATCTCTATACCATTTCTTTAAACGATTTTATG 120  
Db 67 AAAATCATTGATTTATTTATAAATCATAGCAAAATTTATCTCGATATTTATTATAGC 126

QY 121 CTCGGTTTATCGAATAACCTAA 143  
Db 127 GTCATATTGATCTAATAAAGAA 149

## RESULT 10

US-10-312-841-1/c  
; Sequence 1, Application US/10312841  
; Publication No. US20030186277A1  
; GENERAL INFORMATION:  
; APPLICANT: Epigenomics AG  
; TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MEC  
; FILE REFERENCE: E01/1208/WO  
; CURRENT APPLICATION NUMBER: US/10/312,841  
; CURRENT FILING DATE: 2002-12-30  
; NUMBER OF SEQ ID NOS: 2  
; SEQ ID NO 1  
; LENGTH: 3673778

```
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; NAME/KEY: unsure
; LOCATION: (3294164)
US-10-312-841-1

Query Match      12.3%; Score 33.2; DB 12; Length 3673778;
Best Local Similarity 59.6%; Pred. No. 5.8e+02;
Matches 56; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 25 TTTTCTTCCCTCTGTTCCGACCGATTCTCAATAAAAAATCATTCATTCATCGAAGTT 84
Db 2983086 TCTTTTCCCTTCTCTCCACTAACAAATTTCAATAAAACAATATTTTACACCCCAATT 2983027

QY 85 CATTCCTATACCATATCTTTAATACGATTTTA 118
Db 2983026 TCATTAAATATAATAATCTTTAATAATAATTAA 2982993

RESULT 11
US-10-204-708-66/c
; Sequence 66, Application US/10204708
; Publication No. US20030141852A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; TITLE OF INVENTION: by Assessing DNA Methylation
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/204,708
; CURRENT FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 66
; LENGTH: 5360
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-66

Query Match      12.2%; Score 33; DB 12; Length 5360;
Best Local Similarity 54.5%; Pred. No. 74;
Matches 66; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 25 TTTTCTTCCCTCTGTTCCGACCGATTCTCAATAAAAAATCATTCATTCATCGAAGTT 84
Db 3657 TTCTCTTACCTTTTCACTTAACATATATCTTAAATAATCTCTTACAAATACATA 3598

QY 85 CATTCCTATACCATATCTTTAATAACGATTTATGCTCCGGTTTATCGAATAACCTAAC 144
Db 3597 AAAATCTTAACCATCTTTATTACACTACATAATACTCCATTTTATACATACCGTAT 3538

QY 145 T 145
Db 3537 T 3537

RESULT 12
US-10-311-455-1255/c
; Sequence 1255, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
```

```
; Sequence 1910, Application US/10311455
; Publication No. US20030143608A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1910
; LENGTH: 5360
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1910

Query Match      12.2%; Score 33; DB 12; Length 5360;
Best Local Similarity 54.5%; Pred. No. 74;
Matches 66; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 25 TTTTCTTCCCTCTGTTCCGACCGATTCTCAATAAAAAATCATTCATTCATCGAAGTT 84
Db 3657 TTCTCTTACCTTTTCACTTAACATATATCTTAAATAATCTCTTACAAATACATA 3598

QY 85 CATTCCTATACCATATCTTTAATAACGATTTATGCTCCGGTTTATCGAATAACCTAAC 144
Db 3597 AAAATCTTAACCATCTTTATTACACTACATAATACTCCATTTTATACATACCGTAT 3538

QY 145 T 145
Db 3537 T 3537

RESULT 13
US-10-311-455-1255/c
; Sequence 1255, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1255
; LENGTH: 7511
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1255

Query Match      12.1%; Score 32.8; DB 12; Length 7511;
Best Local Similarity 46.8%; Pred. No. 96;
```

```
Matches 103; Conservative 0; Mismatches 117; Indels 0; Gaps 0;
QY 21 CTGTGTTTCTCCCTTCGTTCGACCGATTCTCAATAAAATCAATGTTTCATCGA 80
Db 1375 CATCATTTTCAACCTTTTACTCTTACATATAATTTAACATAAATATCAATCTACTAAA 1316
QY 81 AGTTCAATCCATATACCAATTTCTTTAATAACGATTTTATGCTCGGTTTATCGAATAACC 140
Db 1315 ACCAACTACATATTTTAAAAACAAAATAAACTATAAACTCTTTATTTAATACAAA 1256
QY 141 TAACTTCCACTTCCGTAGCACATGCTAGGATTCGTTATCACTCGGCAATCGCAG 200
Db 1255 AAARACCAATACATAACTCAGGCTATAATCCATCACTTTAAAAAACCAATTTAA 1196
QY 201 GAACAGTGTGGAATACATCTTTACACCCAAATGTTCA 240
Db 1195 AAATCGTCGAACCAAAATCTCGAATCCAACTTTACAA 1156

RESULT 14
US-09-938-842A-4571
; Sequence 4571, Application US/09938842A
; Parent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRI1300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 4571
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-4571

Query Match 12.1%; Score 32.6; DB 10; Length 2000;
Best Local Similarity 54.6%; Pred. No. 63;
Matches 65; Conservative 0; Mismatches 54; Indels 0; Gaps 0;
QY 20 GCTGTTTTTCTCCCTTCGTTCGACCGATTCTCAATAAAATCAATGTTTCATCG 79
Db 766 GTTAATATATATTTCTCCCTCTCTACGTGTTTCAACATATTTCAATTTCTCT 825
QY 80 AGTTTCATTCCTATACCAATTTCTTTAATAACGATTTTATGCTCGGTTTATCGAATAA 138
Db 826 ATTTTCATTCGGGAATAGATCTTTAATCAATGTCATCTCATCTGTCACCAAGTAA 884

RESULT 15
US-10-311-455-1239/c
; Sequence 1239, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
; FILE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
```

```
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1239
; LENGTH: 5218
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1239

Query Match 12.1%; Score 32.6; DB 12; Length 5218;
Best Local Similarity 57.3%; Pred. No. 94;
Matches 59; Conservative 0; Mismatches 44; Indels 0; Gaps 0;
QY 36 TTCTGTTCCGACCGATTCTCAAAATAAAATCAATGATTTCATCGAAGTTTCATCTCTATAC 95
Db 3449 TTAATTAATACTCAATCAAAAATACAAAACAATAATTTAATTAATTAATTCATATAT 3390
QY 96 CATTATCTTTAATAACGATTTTATGCTCCGGTTTATCGAATAA 138
Db 3389 CATATATTTAATTTTAAATTAATAAAATATCCAAAA 3347

Search completed: November 15, 2003, 08:32:16
Job time : 164.996 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: November 14, 2003, 23:56:24 ; Search time 1181.92 Seconds  
(without alignments)  
9241.609 Million cell updates/sec  
Title: US-09-928-457-89  
Perfect score: 267  
Sequence: 1 AATTATGAACACACGATCA.....TTACCACTTCGGATCAATT 267

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 2888711 seqs, 20454813386 residues  
Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*  
1: gb\_ba.\*  
2: gb\_hg.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_on.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*  
29: em\_vi.\*  
30: em\_htg\_hum.\*  
31: em\_htg\_inv.\*  
32: em\_htg\_other.\*  
33: em\_htg\_mus.\*  
34: em\_htg\_pln.\*  
35: em\_htg\_rod.\*  
36: em\_htg\_mam.\*  
37: em\_htg\_vrt.\*  
38: em\_sy.\*  
39: em\_htgo\_hum.\*  
40: em\_htgo\_mus.\*  
41: em\_htgo\_other.\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	267	100.0	267	1	AF169465 Neisseria
2	267	100.0	267	6	A68918 Sequence 89
3	267	100.0	267	6	BD063018 DNA and s
4	241.4	90.4	349061	1	NMA2322491
5	238.2	89.2	10522	1	AE002542 Neisseria
6	238.2	89.2	349980	6	AX044034 Sequence
7	235.8	88.3	339	6	AX024066 Sequence
8	235.8	88.3	339	6	AX033695 Sequence
9	36.8	13.8	145242	9	AC008610 Homo sapi
10	36.8	13.8	178710	2	AC106766 Homo sapi
11	35.6	13.3	178323	2	AC126315 Rattus no
12	35.6	13.3	233045	2	AC132725 Rattus no
13	34.8	13.0	272142	2	AC106944 Rattus no
14	34.6	13.0	827	6	BD147361 Primer fo
15	34.6	13.0	1796	6	AR140098 Sequence
16	34.6	13.0	1845	6	AR140089 Sequence
17	34.6	13.0	1845	6	E15865 cDNA encodi
18	34.6	13.0	1966	6	BD156065 Primer fo
19	34.6	13.0	1966	9	AK001242 Homo sapi
20	34.6	13.0	2046	6	AR140090 Sequence
21	34.6	13.0	2046	6	E15866 cDNA encodi
22	34.6	13.0	3014	9	AB016789 Homo sapi
23	34.6	13.0	3024	6	BD157723 Primer fo
24	34.6	13.0	3024	9	AK022507 Homo sapi
25	34.6	13.0	131355	2	AC139655 Rattus no
26	34.6	13.0	254981	10	AL606479 Mouse DNA
27	34.4	12.9	2000	6	AX655393 Sequence
28	34.2	12.8	11532	1	AE005744 Caulobact
29	33.8	12.7	125020	9	AF429315 Homo sapi
30	33.6	12.6	238046	2	AC095003 Rattus no
31	33.6	12.6	244814	2	AC098223 Rattus no
32	33.4	12.5	174054	2	AC015160 Drosophil
33	33.4	12.5	185497	3	AC010563 Rattus no
34	33.4	12.5	197946	10	AC098722 Mus muscu
35	33.4	12.5	201050	1	AE646084 Ralstonia
36	33.4	12.5	279088	3	AE003518 Drosophil
37	33.2	12.4	906	1	AB072575 Bradyrhiz
38	33.2	12.4	235918	2	AC127741 Rattus no
39	33.2	12.4	258326	2	AC136583 Rattus no
40	33	12.4	645	9	AB018490 Homo sapi
41	33	12.4	3062	9	BC000012 Homo sapi
42	32.8	12.3	136212	2	AC139386 Rattus no
43	32.6	12.2	14188	1	AE007237 Sinorhizo
44	32.6	12.2	172993	9	CNS01RH9 Human chr
45	32.4	12.1	977	8	AF241266 Cucumis m

ALIGNMENTS

RESULT 1  
AF169465  
LOCUS AF169465  
DEFINITION Neisseria meningitidis strain 22491 clone Em029 unknown sequence.  
ACCESSION AF169465  
VERSION AF169465.1 GI:9754673  
KEYWORDS  
ORGANISM Neisseria meningitidis  
SOURCE Neisseria meningitidis  
Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
Neisseriaceae; Neisseria.  
1 (bases 1 to 267)  
REFERENCE  
Perrin, A., Nassif, X. and Tinsley, C.R.  
TITLE Identification of regions of the chromosome of Neisseria meningitidis and Neisseria gonorrhoeae which are specific to

Pred. No. is the number of results predicted by chance to have a



```

QY 121 CTTGGCGGTACCAATGCCGTAAATAATCAGCAACCGCAGCAATGAAGCATACGCATCAA 180
Db 121 CTTGGCGGTACCAATGCCGTAAATAATCAGCAACCGCAGCAATGAAGCATACGCATCAA 180
QY 181 CTTTACCGCAATGTGGGTAAACCGCGGTACCAATGCTATGTTACCAAGTGTATCAGCAC 240
Db 181 CTTTACCGCAATGTGGGTAAACCGCGGTACCAATGCTATGTTACCAAGTGTATCAGCAC 240
QY 241 ATCGGCGTTACCACTTCGATGCAATT 267
Db 241 ATCGGCGTTACCACTTCGATGCAATT 267

RESULT 4
LOCUS NMA222491 349061 bp DNA linear BCT 02-SEP-2002
DEFINITION Neisseria meningitidis serogroup A strain Z2491 complete genome;
ACCESSION AL162753
VERSION AL162753.2 GI:7379120
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS Parkhill,J., Achtman,M., James,K.D., Bentley,S.D., Churcher,C.,
Klee,S.R., Morelli,G., Basham,D., Brown,D., Chillingworth,T.,
Davies,R.M., Davis,P., Devlin,K., Feltwell,T., Hamlin,N.,
Holroyd,S., Jagels,K., Leather,S., Moule,S., Mungall,K.,
Quail,M.A., Rajandream,M.A., Rutherford,K.M., Simmonds,M.,
Skelton,J., Whitehead,S., Spratt,B.G. and Barrall,B.G.
TITLE Complete DNA sequence of a serogroup A strain of Neisseria
meningitidis Z2491
JOURNAL Nature 404 (6777), 502-506 (2000)
MEDLINE 20222556
PUBMED 10761919
REFERENCE
PUBMED 10761919
AUTHORS Parkhill,J.
TITLE Direct Submission
JOURNAL Submitted (30-MAR-2000) Submitted on behalf of the Neisseria
sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk
COMMENTS
Notes:
Details of N. meningitidis sequencing at the Sanger Centre are
available on the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/N_meningitidis/).
FEATURES
source
1. 349061
/organism="Neisseria meningitidis Z2491"
/mol_type="genomic DNA"
/strain="Z2491"
/db_xref="taxon:122587"
/notes="serogroup: A"
complement(24..206)
/notes="ATR repeat; hmms hit to HMW ATR (1 - 183), score:
310.39"
/label="ATR"
209..212
220..681
/genes="NMA0368"
220..681
/genes="NMA0368"
/notes="NMA0368, probable integral membrane protein, len:
153 aa; contains four probable transmembrane domains"
/transl_table=1
/product="putative integral membrane protein"
/protein_id="CAB83669.1"
/db_xref="GI:7379121"
/db_xref="SPTREMBL:Q9JWH7"
/translation="MQEQNRKSPPIVNLVSLVALTASLSNVAFVIGNHGSMEGLTV

```

```

LILGSIASLDIRYCAYVANYVLAAILVLLALEKKVVPVHAAPFWGLALVAFSVKAVYV
DEAGNTSDIVRYGAGFYLWAFVAFVAGISGTFAGKNKERRKAASADGTDKNDV"
671..1492
/genes="hemK"
671..1492
/genes="hemK"
/notes="NMA0369, hemK, Hemk protein, len: 273 aa; similar
to e.g. HEMK_ECOLI P37186 HEMK protein (277 aa), fasta
scores; E(): 0, 42.3% identity in 279 aa overlap. Contains
PS00092 N-6 Adenine-specific DNA methylases signature"
/codon_start=1
/transl_table=1
/product="HemK protein"
/protein_id="CAB83670.1"
/db_xref="GI:7379122"
/db_xref="SPTREMBL:Q9JWH6"
/translation="MTFDKWLGSKLQKNEARMLQYVSEYTRVQLLTRGGEBMPDEV
GRWDLAORLLNGEYVILGARFEYGRFTVPSVLTPRPETELHVAFLARLPEN
RQVADLTGSGAVATVALERPDFAFVRSIDIPPALETARKNAADLCARVEAYGSGWF
DTMPSEGGKWDIIIVSNPPYIENGDKHLSQGLRFEPOIALTDFSDGLSCIRTLAQCAP
DRLAEGGFULLHEHGFQCGAAVRGVLAEFGSVETLPDLAGLDRVTLGKYMHLK"
831..840
/genes="hemK"
/notes="Core DNA uptake sequence: gcgcgtctgaa"
/label=DUS
1162..1171
/genes="hemK"
/notes="Core DNA uptake sequence: gcgcgtctgaa"
/label=DUS
1187..1207
/genes="hemK"
/notes="PS00092 N-6 Adenine-specific DNA methylases
signature"
complement(1279..1288)
/notes="Core DNA uptake sequence: gcgcgtctgaa"
/label=DUS
1565..1568
1578..2966
/genes="NMA0370"
1578..2966
/genes="NMA0370"
/notes="NMA0370, probable integral membrane protein, len:
462 aa; similar to hypothetical proteins e.g. Y325_HAEIN
P4640 hypothetical protein H10325 (450 aa), fasta scores;
E(): 0, 49.8% identity in 464 aa overlap"
/codon_start=1
/transl_table=1
/product="putative integral membrane protein"
/protein_id="CAB83671.1"
/db_xref="GI:7379123"
/db_xref="SPTREMBL:Q9JWH5"
/translation="MNAVVAIVVIMLVLSRVHVVLSTIGAFVGVAGVAGMPLQNIAR
DAAGVQSQAGIIPVFNKLEGKALSYAMLCAPAMATHSGLPQOLAGAVVRKLNAR
GGMPDSVRSGEGVKKWLLLSITLVGMWSONIPIPIAFIMVPLILVFNELKIDR
RLIACVITEGLVTTYWFLPYGGAJFLNEILLGNTHSRAPOLDKXNNVMAAIPAL
GMLAGLLAVHYRKRPRLYQSNNAADTANAGNRRPQPSAYKSAVAIVCFALOL
MTEDSLVGLMGLFAVFMVLGVNDRKNDKADVGEKIKMAWGVFIMIAQQGFVAWNA
TGHIOPLVSSMAIFGNSKGMALMLVVLGVLTWIGTSGSFSTLPIIAIYVPLCVGL
GSPILATVAIVGTAGALDAGSPASDSTLGTPTMGLNADQCHDHSDSVIPTIYNIP
LLIAGWIAAMVL"
2967..3191
/genes="slyX"
2967..3191
/genes="slyX"
/notes="NMA0371, slyX, SLX protein homolog, len: 74 aa;
similar to SLX_HABIN P44759 SLX protein homolog (73 aa),
fasta scores; E(): 0.0017, 33.8% identity in 74 aa
overlap, and SLX_ECOLI P30857 SLX protein (72 aa), fasta
scores; E(): 0.24, 32.4% identity in 68 aa overlap"
/codon_start=1
/transl_table=1
/product="SLX protein homolog"
/protein_id="CAB83672.1"

```

/db\_xref="GI:7379124"  
/db\_xref="SPRMBL:Q9JWH4"  
/translation="MDAQBFHRIITELEISQALQEDVINGNAMVAELRQILDLOQA  
QLRLYQKMDRNPDAQEPYSRLRDEIPPHY"  
complement (3271..3423)  
/gene="NMA0372"  
/complement (3271..3423)  
/gene="NMA0372"  
/translation="unknown, questionable CDS, len: 50 aa"  
/codon\_start=1  
/transl\_table=11  
/product="very hypothetical protein NMA0372"  
/protein\_id="CAB83673.1"  
/db\_xref="GI:7379125"  
/db\_xref="SPRMBL:Q9JWH3"  
/translation="MFRLSLPLSGDGFLLAVYEPVFCICLOGKTAVFRSGGRFHRI  
KGIUSD"  
3562..3571  
/note="Core DNA uptake sequence: gccgtctgaa"  
/label=DUS  
complement (3586..4356)  
/gene="thiP"  
complement (3586..4356)  
/gene="thiP"  
/note="NMA0373, thiF, probable ThiF protein, len: 256 aa,  
similar to e.g. THIF\_ECOLI P30138 ThiF protein (251 aa),  
fasta scores; E(): 0.43.1% identity in 246 aa overlap,  
and MOEB\_ECOLI P12282 molybdopterin biosynthesis MOEB  
protein\_ (249 aa), fasta scores; E(): 0.43.9% identity in  
244 aa overlap (note that N.m. does not have orthologs of  
any other molybdopterin biosynthesis proteins). Contains  
Pfam match to entry PF00899 ThiF\_family, ThiF family"  
/codon\_start=1  
/transl\_table=11  
/product="ThiF protein"  
/protein\_id="CAB83674.1"  
/db\_xref="GI:7379126"  
/db\_xref="SPRMBL:Q9JWH2"  
/translation="MTTEHDNDADFLRYSRHLLEIEGQOKLSAHLIIVGCG  
GTAAALPYLAASIGITLADSDVLEHLRQVAFEDGVGKLTALADLRHIN  
HTVVRTNEKLDGRLTGLVQAADI VLDCDNVATQAVNRACVQAKTPLVSGAAR  
FEGOLAVRPLDPSPCVACLFQDGSASDGLCSLFGVPSLVGIIQSTQAAEALKIL  
DAGSPSGRLAVYALEGQWYFLPRNPECPVCAER"  
3787..3796  
/note="Core DNA uptake sequence: gccgtctgaa"  
/label=DUS  
complement (3847..4257)  
/gene="thiF"  
/note="Pfam match to entry PF00899 ThiF\_family, ThiF  
family, score 186.60, E-value 4.1e-52"  
complement (4418..4427)  
/note="Core DNA uptake sequence: gccgtctgaa"  
/label=DUS  
4471..7224  
/gene="ppc"  
4471..7224  
/gene="ppc"  
/ec\_number="4.1.1.31"  
/note="NMA0374, ppc, phosphoenolpyruvate carboxylase, len:  
917 aa; similar to many e.g. CAPP\_RHOFA Q32483  
phosphoenolpyruvate carboxylase (EC 4.1.1.31) (936 aa),  
fasta scores; E(): 0.43.3% identity in 928 aa overlap.  
Contains 2x Pfam match to entry PF00311 PEPCase,  
Phosphoenolpyruvate carboxylase, PS00017 ATP/GTP-binding  
site motif A (P-loop), and PS00393 Phosphoenolpyruvate  
carboxylase active site 2"  
/codon\_start=1  
/transl\_table=11  
/product="phosphoenolpyruvate carboxylase"  
/protein\_id="CAB83675.1"

Query Match 90.4%; Score 241.4; DB 1; Length 349061;  
Best Local Similarity 98.2%; Pred. No. 1.1e-59;

Matches 266; Conservative 0; Mismatches 1; Indels 4; Gaps 2;  
QY 1 AATTATGAACACACGATCATCGTTTCGGCTGCGTTTCGTTGGCATTAGCAGGTTG 60  
Db 149147 AATTATGAACACACGATCATCGTTTCGGCTGCGTTTCGTTGGCATTAGCAGGTTG 149206  
QY 61 CGGCTCAATCAATATGTAACCGTTTCGACACGAAATCTTCAGGAACGTCGCGGTTTGC 120  
Db 149207 CGGCTCAATCAATATGTAACCGTTTCGACACGAAATCTTCAGGAACGTCGCGGTTTGC 149266  
QY 121 CTTGGCGTCA---CCAATGCGGTAAATCAGAACCGCAGCAATGAAGCATACGCAT 177  
Db 149267 CTTGGCGTCAAGCCCAATGCGGTAAATCAGAACCGCAGCAATGAAGCATACGCAT 149326  
QY 178 CAACATTTACCGCAACTGTGGTAAAGCGGTGA-CAATGCTATGTTACCATGTAATCAG 236  
Db 149327 CAACATTTACCGCAACTGTGGTAAAGCGGTGAAGCCATGCTATGTTACCATGTAATCAG 149386  
QY 237 CACAATGCGCGTTACCACTTCGGATGCAATT 267  
Db 149387 CACAATGCGCGTTACCACTTCGGATGCAATT 149417

RESULT 5  
AE002542/c  
LOCUS  
DEFINITION  
Neisseria meningitidis serogroup B strain MC58 section 184 of 206  
of the complete genome.  
ACCESSION  
AE002542 AE002098  
VERSION  
AE002542.1 GI:7227188  
KEYWORDS  
Neisseria meningitidis MC58  
SOURCE  
Neisseria meningitidis MC58  
ORGANISM  
Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
Neisseriaceae; Neisseria.  
REFERENCE  
1 (bases 1 to 10522)  
Tettelin, H., Saunders, N.J., Heidelberg, J., Jeffries, A.C.,  
Nelson, K.E., Eisen, J.A., Ketchum, K.A., Hood, D.W., Pedersen, J.F.,  
Dodson, R.J., Nelson, W.C., Gwinn, M.L., DeBoy, R., Peterson, J.D.,  
Hickey, E.K., Haft, D.H., Salzberg, S.L., White, O., Fleischmann, R.D.,  
Dougherty, B.A., Mason, T., Ciecko, A., Parksey, D.S., Blair, E.,  
Cittone, H., Clark, E.B., Cotton, M.D., Utterback, T.R., Khouri, H.,  
Qin, H., Vamathevan, J., Gill, J., Scarlato, V., Masignani, V.,  
Pizza, M., Grandi, G., Sun, L., Smith, H.O., Fraser, C.M., Moxon, E.R.,  
Rappuoli, R. and Venter, J.C.  
Complete genome sequence of Neisseria meningitidis serogroup B  
strain MC58  
Science 287 (5459), 1809-1815 (2000)  
JOURNAL  
MEDLINE  
20175755  
10710307  
PUBMED  
REFERENCE  
2 (bases 1 to 10522)  
Tettelin, H., Saunders, N.J., Heidelberg, J., Jeffries, A.C.,  
Nelson, K.E., Eisen, J.A., Ketchum, K.A., Hood, D.W., Pedersen, J.F.,  
Dodson, R.J., Nelson, W.C., Gwinn, M.L., DeBoy, R., Peterson, J.D.,  
Hickey, E.K., Haft, D.H., Salzberg, S.L., White, O., Fleischmann, R.D.,  
Dougherty, B.A., Mason, T., Ciecko, A., Parksey, D.S., Blair, E.,  
Cittone, H., Clark, E.B., Cotton, M.D., Utterback, T.R., Khouri, H.,  
Qin, H., Vamathevan, J., Gill, J., Scarlato, V., Masignani, V.,  
Pizza, M., Grandi, G., Sun, L., Smith, H.O., Fraser, C.M., Moxon, E.R.,  
Rappuoli, R. and Venter, J.C.  
Direct Submission  
Submitted (17-MAR-2000) The Institute for Genomic Research, 9712  
Medical Center Dr, Rockville, MD 20850, USA  
Location/Qualifiers  
1..10522  
/organism="Neisseria meningitidis MC58"  
/mol\_type="genomic DNA"  
/strain="MC58"  
/db\_xref="taxon:122586"  
/note="serogroup: B"  
complement (72..419)  
/gene="NMB1931"  
complement (72..419)

TITLE  
JOURNAL  
source  
FEATURES  
source  
TITLE  
JOURNAL  
Submitted (17-MAR-2000) The Institute for Genomic Research, 9712  
Medical Center Dr, Rockville, MD 20850, USA  
Location/Qualifiers  
1..10522  
/organism="Neisseria meningitidis MC58"  
/mol\_type="genomic DNA"  
/strain="MC58"  
/db\_xref="taxon:122586"  
/note="serogroup: B"  
complement (72..419)  
/gene="NMB1931"  
complement (72..419)

gene  
CDS



```

/gene="NMB1931"
/notes="hypothetical protein; identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF42260.1"
/db_xref="GI:7227189"
/transl_table="MPVYFRTPSRKRLHKNRRQKHLGLGEFOTLIFGVGSLHQAE
DAAEFDRFMELIAYLAANGFEMAGGLRRFFVFSKIRCPKLRNREKSHYLLAGR
PRRCHFVRHPAG"
/complement(496..1401)
/gene="NMB1932"
/complement(496..1401)
/gene="NMB1932"
/notes="similar to SP:P00960 GB:J01622 PID:146222
PID:146225 PID:466698 percent identity: 85.67; identified
by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="Glycyl-tRNA synthetase, alpha chain"
/protein_id="AAF42261.1"
/db_xref="GI:7227190"
/transl_table="MLTFOQIIFKLTFWADKGTCLQPPDMVEGAGTSHPATCLRAL
GPEPFAAYVOPSRPKDGRYQDNPNLQHYVQFVALKPPANIQIYLDLSRELGI
DPKVDIIEFVEDDRENPTLGAAGLQWELNGMEVTFYFQQVGGIDCFVLGHEIT
GIERLAMYQLGVENYIDLWAKTLOGNTVTYGDVHQNEVEQSYNFEYSDADWLLRQ
FNDEAQAQRLLAEENALALPAYELVLKAGHTNLLDARGAISVTERATYIGIRAL
SRAVQXYVESREKLGFFPMKANAA"
/complement(1806..2228)
/gene="NMB1933"
/complement(1806..2228)
/gene="NMB1933"
/notes="similar to GB:LI0328 SP:P00832 GB:M25464 GB:V00267
GB:V00311 percent identity: 66.92; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="ATP synthase F1, epsilon subunit"
/protein_id="AAF42262.1"
/db_xref="GI:7227191"
/transl_table="MSINQVEVSESEQIXSGEATFIVVPTQVQELGIYPRHPEIMSL
VRGALATVPEGEKEVLVAVSGGVLEVPQDKVTLADVAVRSAEMDRARAEEAKAA
EAGISQAQKDDKALAEAHKVLAAIAQLKTLDIYRSKK"
/complement(2239..3636)
/gene="NMB1934"
/complement(2239..3636)
/gene="NMB1934"
/notes="similar to SP:P42468 PID:587502 percent identity:
91.52; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="ATP synthase F1, beta subunit"
/protein_id="AAF42263.1"
/db_xref="GI:7227192"
/transl_table="MSQGIQVIIGAVDVVEFRDMPRVYDALKILDENGLITILEVOQL
LGDGVRAIAMSGSDGLKRGMTVNTGAPITVPVKGTLGRIVDVLGTVPVDEAGPDT
DKSRAIHOAPKFDLSSTTELTETGIKVIDLLCPRAKGGKVLFGGAGVGTVMNME
LNNIARAHAGSLSVFAGVGRTEBGNDFYHEMKDSNVLVDKVMYCGMPPGPNRLAV
ALNTGLTAEVPRDEKNGRQDVLFFVDNIYRTLAGTEVSALLGMPSSAVGQPTL
AENRQCRITSTQTSITSIQAVYVPAADDLTPSPATTFADLDTATVLSRDIASIG
IYPAVDLDDSTQSLDQVQHEHYDVARGVQSTLQKYKELRDIILILGMDLSDDEDK
LTVNRARKIQRFLSQPFVHAVEVETGSPGKYVALRDTIAGFKAILNGEYDHLPEQAFYM
VGSIEEAVEKATLN"
/complement(3674..4549)
/gene="NMB1935"
/complement(3674..4549)
/gene="NMB1935"
/notes="similar to GB:LI0328 SP:P00837 GB:M25464 GB:V00267
PID:146324 percent identity: 75.43; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11

```

```

/product="ATP synthase F1, gamma subunit"
/protein_id="AAF42264.1"
/db_xref="GI:7227193"
/transl_table="MAVGKEILTKIRSVQNTQKITRAMQVSTSKMRKTQERWRLARP
YAEKVRWWSHLAQTNIDHGIPLLESHREIRRVGFILITSDKGLCGGLNANVLFKFLA
QVQYRNQIGIEEIVFCGSKGLMACQSIGLVNVASVNLGDTPKMELMLLCPTELFOR
YEKHEIDRIHLVYGVFWNTMRQPRMVELLPIGENVIGDSAPKSPSWERYEPTTALA
ITTELRYRLEYSVVYQALSDNMAEQAARWAMKATDNAGNAIKELRLVYNKSRQAA
LTVLELYVAGAAV"
/complement(4574..6121)
/gene="NMB1936"
/complement(4574..6121)
/gene="NMB1936"
/notes="similar to SP:P12985 PID:48337 percent identity:
85.77; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="ATP synthase F1, alpha subunit"
/protein_id="AAF42265.1"
/db_xref="GI:7227194"
/transl_table="MQLNPAEISDLIKAKIENLSVNAEVRTCQTVISVTDGIVRIHGL
SDAMQEMLEFPNGTETGLAMNLERDSVGAIVLGEVEHIEKGGDTVCTGRILEVPGRE
LVGRVVDALGRPIDGKGPINTLTAPIEKIAIGVITARKSVDPQPMOTGLKAISMPVPG
RGRRELIIGDROGTGTAVALDAIVNKGIGVICIYVAICQKASSANVRKUEEHGAM
EHTIVVAATASAAALQYIAPYSGCTMGFEFFDRGEDALIVYDDLSKQAVAYRQISLL
LRPPREAYPGDVFYLRHLLERAAARVNEHEVETLNGEVKGTGSITALPIIETQA
GDVSAEVPNTVISITDQIFLETDLFNAGIRPAINAGISVSRVGGAAQTKVKKLGGQ
IRLALAQYELAAFSQFASDLDEATRKQLEHGEVETELMKQKQFSTLTAEMALTWA
INNGSYSDYPVAKALAFSEFSLFVTRQHPVLEAVNASGMSDESEKLEAMKSFK
SVAYQA"
/complement(6132..6665)
/gene="NMB1937"
/complement(6132..6665)
/gene="NMB1937"
/notes="similar to SP:P12987 PID:48336 percent identity:
59.30; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="ATP synthase F1, delta subunit"
/protein_id="AAF42266.1"
/db_xref="GI:7227195"
/transl_table="NAEPATIAAPYAKALFGLAQEKNOIESWLGLEKLAAYVQBGKV
ASLDIREFENASEKADILDLVGLDKELKNEVILVLAGOKRLILPEVYAOQCDLTLS
FHLIKSAVIYSAYPLTDRQVGLVQMLNKRFRFSELKISVEIEPELIGGKKEVGGQOVL
DLSVQKUSALYITMTN"
/complement(6670..7140)
/gene="NMB1938"
/complement(6670..7140)
/gene="NMB1938"
/notes="similar to SP:P12989 PID:581814 percent identity:
68.59; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="ATP synthase F0, B subunit"
/protein_id="AAF42267.1"
/db_xref="GI:7227196"
/transl_table="WNINATLFAOIIIVFGLVFTMKFWPPIAKALDERAAKVAEGL
AAARSGSDFEQAEKKVALLAEQNSVWVAEKRRAKIVVEAKQASSEAAARIA
AQAKADVQELPRAREUREQVAVLAVGAESILRSEVDASHKXLLDTLQOEL"
/complement(7211..7447)
/gene="NMB1939"
/complement(7211..7447)
/gene="NMB1939"
/notes="similar to SP:P12991 PID:48334 percent identity:
73.75; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="ATP synthase F0, C subunit"
/protein_id="AAF42268.1"
/db_xref="GI:7227197"
/transl_table="MGLTAAICCLIVLALGASIGIANVGSKYLESSARQPELIGPL
QTKFLIAGLIDAAFLIGLVAIALLFVFNPPAG"
/complement(7504..8370)

```

Gene



```

RESULT 8
AX033695
LOCUS      AX033695          339 bp    DNA          linear          PAT 21-SEP-2000
DEFINITION Sequence 1 from Patent WO0044904.
ACCESSION  AX033695
VERSION     AX033695.1    GI:10280379
KEYWORDS   Neisseria meningitidis
SOURCE      Neisseria meningitidis
ORGANISM   Neisseria meningitidis
            Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
            Neisseriaceae; Neisseria.
            Thonard,J.
REFERENCE   Bas059 polypeptides from reisseria meningitidis
AUTHORS     Smithline Beecham Biolog (BE) ; THONNARD JOELLE (BE)
JOURNAL     SMITHKLINE BEECHAM BIOLOG (BE)
FEATURES   Location/Qualifiers
            1..339
            /organism="Neisseria meningitidis"
            /mol_type="genomic DNA"
            /db_xref="taxon:487"
            93 a      84 c      78 t
BASE COUNT  93 a      84 c      78 t
ORIGIN
Query Match      88.3%; Score 235.8; DB 6; Length 339;
Best Local Similarity 97.8%; Pred. No. 4e-59;
Matches 261; Conservative 0; Mismatches 2; Indels 4; Gaps 2;
QY      5 ATGACACACAGCATCATGTTTCGGCTGCGTTCGTTGGTGGCATTAGCAGGTTGGGC 64
Db      1 ATGACACACAGCATCATGTTTCGGCTGCGTTCGTTGGCATTAGCAGGTTGGGC 60
QY      65 TCAATCAATAATGTAACCGTTTCGACCCAGAACTTCAGGAACGTCGCGTTGCGCTTG 124
Db      61 TCAATCAATAATGTAACCGTTTCGACCCAGAACTTCAGGAACGTCGCGTTGCGCTTG 120
QY      125 GCGGTCACCC---AATGCGGTAAATATCAGCACCGCAGCAATGAGGCATACGATCAAC 181
Db      121 GCGGTCAGCCAAATGCGGTAAATATCAGCACCGCAGCAATGAGGCATACGATCAAC 180
QY      182 TTACCGCAACTGTGGTGAAGCGGTGA-CCAATGCTATGTTACAGTGTAAATCAGCACA 240
Db      181 TTACCGCAACTGTGGTGAAGCGGTGAAGCAATGCTATGTTACAGTGTAAATCAGCACA 240
QY      241 ATCGCGGTACCACTTCGATGCAATT 267
Db      241 ATCGCGGTACCACTTCGATGCAATT 267

RESULT 9
AC008610/c
LOCUS      AC008610          145242 bp    DNA          linear          PRI 15-AUG-2001
DEFINITION Homo sapiens chromosome 5 clone CTB-12904, complete sequence.
ACCESSION  AC008610
VERSION     AC008610.7    GI:15187193
KEYWORDS   HTG.
SOURCE      Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 145242)
AUTHORS     DOE Joint Genome Institute and Stanford Human Genome Center.
JOURNAL     Direct Submission
TITLE       Direct Submission
REFERENCE   2 (bases 1 to 145242)
AUTHORS     DOE Joint Genome Institute.
JOURNAL     Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3 (bases 1 to 145242)
AUTHORS     DOE Joint Genome Institute and Stanford Human Genome Center.

Direct Submission
Submitted (25-JAN-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
4 (bases 1 to 145242)
DOE Joint Genome Institute and Stanford Human Genome Center.
Direct Submission
Submitted (15-AUG-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
On Aug 15, 2001 this sequence version replaced gi:12484308.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.sngc.stanford.edu
Quality: Phrap Quality >=40 99.6% of Sequence;
Estimated Total Number of Errors is 0.5.
STS Content:
WI-2865 G04046.
Location/Qualifiers
1..145242
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTB-12904"
37848 a 36059 c 36278 g 35057 t
BASE COUNT 37848 a 36059 c 36278 g 35057 t
ORIGIN
Query Match      13.8%; Score 36.8; DB 9; Length 145242;
Best Local Similarity 55.5%; Pred. No. 11;
Matches 71; Conservative 0; Mismatches 57; Indels 0; Gaps 0;
QY      100 TCAGGAACGTCCGCGCTTTGCTTGGCGGTCCACCAATGCGTAAAAATACGACACGCAG 159
Db      62414 TAAGGACCGGCGGCTCTCACCGTGGCGTCCACCAACCGTGGGAGCTCCATCTCTCG 62355
QY      160 CAATGAAGGCATACGCATCAACTTTACCGCACTGTGGGTAAAGCGGTGACCAATGCTAT 219
Db      62354 CGAGACCGACTGCGGCGTCCACATCAACGACGCGCGAGATCGGTGGCAGCACCA 62295
QY      220 GTTACCAG 227
Db      62294 GTTAGGAG 62287

AC106766      178710 bp    DNA          linear          HTG 12-JAN-2002
LOCUS      AC106766/c
DEFINITION Homo sapiens chromosome 5 clone RP11-252M5, WORKING DRAFT SEQUENCE,
            11 unordered pieces.
ACCESSION  AC106766
VERSION     AC106766.1    GI:18139316
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 178710)
AUTHORS     DOE Joint Genome Institute.
JOURNAL     Sequencing of Human Chromosome 5
TITLE       Unpublished
REFERENCE   2 (bases 1 to 178710)
AUTHORS     DOE Joint Genome Institute.
JOURNAL     Direct Submission
TITLE       Direct Submission
JOURNAL     Submitted (12-JAN-2002) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     -----Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -----
            Project Information
            Center Project Name: 503426
            Center clone name: RPC1-11_252M5

```

-----  
 Summary Statistics  
 Consensus quality: 168617 bases at least Q40  
 Consensus quality: 170711 bases at least Q30  
 Consensus quality: 171417 bases at least Q20  
 Estimated insert size: 176230; agarose-fp estimation  
 Estimated insert size: 177710; sum-of-contigs estimation  
 Quality coverage: 10.04 in Q20 bases; agarose-fp estimation  
 Quality coverage: 10.07 in Q20 bases; sum-of-contigs estimation.  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 11 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

\* 1712: contig of 1712 bp in length  
 \* 1713: gap of unknown length  
 \* 1812: gap of unknown length  
 \* 1813: contig of 1822 bp in length  
 \* 3504: gap of unknown length  
 \* 3505: gap of unknown length  
 \* 4916: contig of 1312 bp in length  
 \* 5016: gap of unknown length  
 \* 53771: contig of 2581 bp in length  
 \* 7597: gap of unknown length  
 \* 7598: gap of unknown length  
 \* 13612: contig of 5915 bp in length  
 \* 13712: gap of unknown length  
 \* 13713: gap of unknown length  
 \* 31775: contig of 18063 bp in length  
 \* 31776: gap of unknown length  
 \* 31875: gap of unknown length  
 \* 53671: contig of 21796 bp in length  
 \* 53672: gap of unknown length  
 \* 53771: contig of 20004 bp in length  
 \* 73775: gap of unknown length  
 \* 73776: gap of unknown length  
 \* 73776: gap of unknown length  
 \* 95412: contig of 21537 bp in length  
 \* 95413: gap of unknown length  
 \* 95513: contig of 30824 bp in length  
 \* 126336: gap of unknown length  
 \* 126337: gap of unknown length  
 \* 126437: 178710: contig of 52274 bp in length.

## FEATURES

## SOURCE

1. 178710  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="5"  
 /clone="RP11-252M5"  
 /clone\_lib="RPCI human BAC library 11"

BASE COUNT 45146 a 44147 c 43668 g 44699 t 1050 others  
 ORIGIN

Query Match 13.8%; Score 36.8; DB 2: Length 178710;  
 Best Local Similarity 55.5%; Pred. No. 11;  
 Matches 71; Conservative 0; Mismatches 57; Indels 0; Gaps 0;  
 QY 100 TCAGGAGTCGCGGTTGCTTGGCGTCCACCAATCCCGTAAATACGACCAACGAG 159  
 DB 127083 TAAGGACCGCGCGTCTCCCGTGGCGTCCACCAACCGCGGAGATCGGCGTCCATCTCTCG 127024  
 QY 160 CAATGAAGGCATACGCATCAACTTACCGCAACTGTGGTAAGCGCGTACCAATGCTAT 219  
 DB 127023 CGAGACCGACTGCGCGTCCACATCAACGCGGCGGAGATCGGCGTCCATCTCTCG 126964  
 QY 220 GTTACCAG 227  
 DB 126963 GTTAGGAG 126956

## RESULT 11

## AC126315

## LOCUS

## DEFINITION

## AC126315

## ACCESSION

## VERSION

## GI:25080517

AC126315 178233 bp DNA linear HTG 19-NOV-2002

Rattus norvegicus clone CH230-409C19, \*\*\* SEQUENCING IN PROGRESS

\*\*\*, 3 unordered pieces.

AC126315

AC126315.4 GI:25080517

KEYWORDS  
SOURCE  
ORGANISM

HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.  
 Rattus norvegicus (Norway rat)

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.

## REFERENCE

## AUTHORS

1 (bases 1 to 178323)

Muzny, D., Marie, J., Mettler, M., Lee, J., Adams, C., Alder, J.,  
 Allen, C., Allen, H., Altschuld, S., Amin, A., Anguiano, D.,  
 Anyalechechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,  
 Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,  
 Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,  
 Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,  
 Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,  
 Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,  
 Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,  
 Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,  
 Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,  
 Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,  
 Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,  
 Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,  
 Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,  
 Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,  
 Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,  
 Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,  
 Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M.,  
 Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,  
 Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,  
 Karpathy, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowalski, J.,  
 Kowalski, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,  
 Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,  
 Lorensu, H., Loui, L., Loui, S., Lozano, R.J., Lu, X., Ma, J.,  
 Maheshwari, M., Mahindartne, M., Mahmoud, M., Mallory, K., Mangum, A.,  
 Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,  
 Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,  
 Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,  
 Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,  
 Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Parks, K.,  
 Nwankwelu, O., Okwionu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,  
 Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankuch, C.,  
 Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.,  
 Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Reiter, M.A., Reigh, R.,  
 Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,  
 Rives, C., Rodkey, T., Rojas, A., Rose, R., Rose, R., Ruiz, S., Shen, H.,  
 Sanders, W., Savary, G., Scherer, S., Scott, G., Shatman, S., Shetty, J.,  
 Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smales, D.,  
 Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,  
 Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,  
 Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umani, K.,  
 Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,  
 Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,  
 Williams, G., Willson, R., Wleciyk, R., Wooden, H., Worley, K.,  
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,  
 Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Zhao, S., Dunn, D., von  
 Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,  
 Weinstein, G. and Gibbs, R.A.

Direct Submission

Unpublished

2 (bases 1 to 178323)

Worley, K.C.

Direct Submission

Submitted (05-JUL-2002)

Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 178323)

Direct Submission

Rat Genome Sequencing Consortium.

Submitted (19-NOV-2002)

Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

On Nov 19, 2002 this sequence version replaced gi:23195448.

The sequence in this assembly is a combination of BAC based reads

and whole genome shotgun sequencing reads assembled using Atlas

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: http://www.hgsc.bcm.tmc.edu/  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GZ08  
Center clone name: CH230-409C19  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329  
Consensus quality: 145050 bases at least Q40  
Consensus quality: 147946 bases at least Q30  
Consensus quality: 149916 bases at least Q20  
Estimated insert size: 150833; sum-of-contigs estimation  
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

-----  
\* NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html).  
\* NOTE: This is a 'working draft' sequence. It currently consists of 3 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

\* 1 171060: contig of 171060 bp in length  
\* 171161 171160: gap of unknown length  
\* 171161 173957: contig of 2797 bp in length  
\* 173958 174057: gap of unknown length  
\* 174058 178323: contig of 4266 bp in length.

FEATURES

source  
1..178323  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-409C19"  
1..1376  
/note="wgs end\_extension"  
clone\_end:17"  
3965..4975  
/note="clone\_boundary"  
clone\_end:17"  
site:  
end\_sequence:B2249410"  
complement(167238..168180)  
/note="clone\_boundary"  
clone\_end:596  
site:  
end\_sequence:RXBBA22TV"  
41076 a 34387 c 34604 g 41395 t 26861 others

BASE COUNT

ORIGIN  
Query Match 13.3%; Score 35.6; DB 2; Length 178323;  
- Best Local Similarity 50.6%; Pred. No. 24;  
Matches 86; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

QY 50 TTACGAGTTGGGTCATCAATGAATGTAACCGTTTCGACGAGAACTTCAGGAACGT 109  
|||||

Db 59751 TTGAATCGTAAAGCAAGATCTTGCTATATCTGTGTTAAACGAGAACTTTAGAAACTT 59810  
|||||

QY 110 GCCCGTTGCTGGCGCTACCAATCCCGTAAATACCAACCCGACCAATGAAGGC 169  
|||||

TITLE  
JOURNAL  
Unpublished  
REFERENCE  
2 (bases 1 to 233045)

Db 59811 TAGAAACTCCCTGGAGCCTTAAGATTCTTCTGTATTATCACCAGCAGACTAGATACGCA 59870  
QY 170 ATACGGATCAACTTACCCGACTGTGGTAAAGCGGTGACCAATGCTAT 219  
|||||

Db 59871 TGTATAGAACCTTTAGACCAATTGGAACACACACTTAAACAAATGCTAT 59920  
|||||

RESULT 12

AC132725 233045 bp DNA linear HTG 19-NOV-2002  
Rattus norvegicus clone CH230-351P24, WORKING DRAFT SEQUENCE, 6  
unordered pieces.  
AC132725

AC132725.3 GI:25089932  
HTG; HTGS PHASE1; HTGS DRAFT; HTGS\_FULLTOP.

Rattus norvegicus (Norway rat)

Rattus norvegicus

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

REFERENCE

AUTHORS

1 (bases 1 to 233045)  
Muzny, D. Marie., Metzker, M. Lee., Abramzon, S., Adams, C., Alder, J.,  
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,  
Anyalebechi, V., Ayodeji, A., Ayodeji, M., Baca, E., Baden, H.,  
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,  
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,  
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,  
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,  
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,  
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,  
Davilla, W. L., Davis, C., Davy-Carroil, L., De Andrada, C., Dederich, D.,  
Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,  
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, X., Duval, B., Eaves, K.,  
Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,  
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,  
Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,  
Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,  
Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,  
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,  
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,  
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,  
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,  
Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,  
Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,  
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,  
Lorenshewa, L., Loulseghe, H., Lozano, R. J., Lu, X., Ma, J.,  
Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,  
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,  
Mawhiney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,  
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,  
Morgan, M., Morris, K., Morris, S., Mundasa, M., Murphy, M., Nait, L.,  
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,  
Nwankwelu, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,  
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,  
Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.,  
Puazo, M., Quitoz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,  
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,  
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J.,  
Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,  
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajls, D.,  
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,  
Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,  
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umani, K.,  
Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,  
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,  
Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K.,  
Wright, J., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,  
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von  
Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,  
Weinstock, G. and Gibbs, R. A.

Direct Submission

Unpublished

2 (bases 1 to 233045)

**AUTHORS** Rat Genome Sequencing Consortium.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (03-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

**REFERENCE** 3 (bases 1 to 233045)  
**AUTHORS** Rat Genome Sequencing Consortium.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

**COMMENT**  
 On Nov 19, 2002 this sequence version replaced gi:23811366.  
 The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: GUQI  
 Center clone name: CH230-351P24  
 ----- Summary Statistics  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 205239 bases at least Q40  
 Consensus quality: 207848 bases at least Q30  
 Consensus quality: 209231 bases at least Q20  
 Estimated insert size: 208878; sum-of-contigs estimation  
 Quality coverage: 7x in Q20 bases; sum-of-contigs estimation  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank/draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html)).  
 \* consists of 6 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 61478: contig of 61478 bp in length  
 61479 61578: gap of unknown length  
 61579 68503: contig of 6925 bp in length  
 68504 68603: gap of unknown length  
 68604 154066: contig of 85463 bp in length  
 154067 154166: gap of unknown length  
 154167 230594: contig of 76428 bp in length  
 230595 231905: gap of unknown length  
 231906 232005: contig of 1211 bp in length  
 232006 233045: contig of 1040 bp in length.

**FEATURES** Location/Qualifiers  
 1..233045  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10116"  
 /clone="CH230-351P24"  
 2676..3550  
 /note="clone boundary  
 clone\_end:5p6"  
 site:  
 end\_sequence:B225033"  
 53991..54868

**misc\_feature**  
 misc\_feature  
 misc\_feature

/note="clone boundary  
 clone\_end:17"  
 end\_sequence:B225031"  
 57920..60327  
 /note="wgs\_end\_extension.  
 clone\_end:17"  
 60378..61478  
 /note="wgs\_end\_extension  
 clone\_end:17"  
 154167..155279  
 /note="wgs\_end\_extension  
 clone\_end:17"  
 58983 a 47651 c 47755 g 55805 t 22851 others  
 ORIGIN  
 Query Match 13.3%; Score 35.6; DB 2; Length 233045;  
 Best Local Similarity 50.16%; Pred. No. 24;  
 Matches 86; Conservative 0; Mismatches 84; Indels 0; Gaps 0;  
 QY 50 TTAGCAGGTTCGGCTCAATCAATGTAACCGTTCCGACCCAGAACTTCAGGAAGT 109  
 Db 10340 TTAGATGTTAGCAAGATCTCTCTATCTGTTTAAACCGAACTTTGAAACTT 10399  
 QY 110 GCGCGTTTTCCTGGCGTCCACCAATCCGTAATAAATCCAGCAACCGCAGCAATGAAGGC 169  
 Db 10400 TAGAAACTTCCTGGAGCCTTAAGATTCTCTGTATACCGACGACACTAGATACGCA 10459  
 QY 170 ATACGATCACTTTACCGCAACTGCGGTAGCGCGTGACCAATGCTAT 219  
 Db 10460 TGTATAGAACCTTTAGACCAATTTGGAACACACCTTAACAAATGCTAT 10509

**RESULT** 13  
**LOCUS** AC106944/c  
**DEFINITION** Rattus norvegicus clone CH230-3718, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 5 unordered pieces.  
**ACCESSION** AC106944  
**VERSION** AC106944.4 GI:22857067  
**KEYWORDS** HTG; HTGS PHASE1; HTGS DRAFT; HTGS\_ENRICHED.  
**SOURCE** Rattus norvegicus (Norway rat)  
**ORGANISM** Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 1 (bases 1 to 272142)  
 Muzny, D. Marie, Metzker, M. Lee, Abramson, S., Adams, C., Alder, J.,  
 Allen, C., Allen, H., Albrooks, S., Amin, A., Anguiano, D.,  
 Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,  
 Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,  
 Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,  
 Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,  
 Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,  
 Chacko, J., Chavez, D., Chen, G., Chen, Y., Chen, Z., Chu, J.,  
 Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,  
 Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,  
 Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,  
 Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,  
 Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,  
 Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,  
 Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,  
 Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, W.,  
 Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,  
 Harvey, J., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,  
 Hernandez, R., Hines, S., Hladun, S. B., Hodgson, A., Hogues, M.,  
 Hollins, B., Howells, S., Huiyk, S., Hume, J., Idlebird, D., Jackson, A.,  
 Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,  
 Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,  
 Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,  
 Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,  
 Lorensuhewa, L., Louissegh, H., Lozano, R. J., Lu, X., Ma, J.,  
 Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,

Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,  
 Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,  
 Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,  
 Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,  
 Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,  
 Nwaokelemeh, O., Okwuon, G., Olarnpunsagoon, A., Pal, S., Parks, K.,  
 Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,  
 Plopper, F., Foidexter, A., Popovic, D., Primus, E., Pu, L.-L.,  
 Puzo, M., Quiroz, J., Rachin, B., Reeves, K., Regier, M.A., Reigh, R.,  
 Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,  
 Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,  
 Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,  
 Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D.,  
 Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,  
 Staimis, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,  
 Taylor, I., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umani, K.,  
 Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,  
 Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,  
 Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K.,  
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,  
 Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von  
 Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,  
 Weinstock, G. and Gibbs, R.A.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 272142)  
 Worley, K.C.  
 Direct Submission  
 Submitted (14-JAN-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 272142)  
 Rat Genome Sequencing Consortium.  
 Direct Submission  
 Submitted (08-OCT-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Sep 14, 2002 this sequence version replaced gi:21737096.  
 The sequence in this assembly is a combination of BAC based reads  
 and whole genome shotgun sequencing reads assembled using Atlas  
 (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described  
 in the feature table below represents a scaffold in the Atlas  
 assembly (a 'contig-scaffold'). Within each contig-scaffold,  
 individual sequence contigs are ordered and oriented, and separated  
 by sized gaps filled with Ns to the estimated size. The sequence  
 may extend beyond the ends of the clone and there may be sequence  
 contigs within a contig-scaffold that consist entirely of whole  
 genome shotgun sequence reads. Both end sequences and whole genome  
 shotgun sequence only contigs will be indicated in the feature  
 table.

----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: http://www.hgsc.bcm.tmc.edu/  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: GONT  
 Center clone name: CH230-3718  
 ----- Summary Statistics  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 220398 bases at least Q40  
 Consensus quality: 224532 bases at least Q30  
 Consensus quality: 226624 bases at least Q20  
 Estimated insert size: 243846; sum-of-contigs estimation  
 Quality coverage: 3x in Q20 bases; sum-of-contigs estimation  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 consists of 5 contigs. The true order of the pieces  
 is not known and their order in this sequence record is  
 arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 264441: contig of 264441 bp in length  
 264442: gap of unknown length  
 264541: contig of 1119 bp in length  
 265660: gap of unknown length  
 265661: gap of unknown length  
 265760: gap of unknown length  
 265761: contig of 1726 bp in length  
 267487: gap of unknown length  
 267586: gap of unknown length  
 267587: contig of 1166 bp in length  
 268753: gap of unknown length  
 268753: gap of unknown length  
 268853: contig of 3290 bp in length.

FEATURES  
 Location/Qualifiers  
 1..272142  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10116"  
 /clone="CH230-3718"  
 1..1619  
 /note="wgs contig"  
 misc\_feature  
 260733..261787  
 /note="wgs contig"  
 misc\_feature  
 71571 a 43698 c 43504 g 69419 t 43950 others  
 BASE COUNT  
 ORIGIN

Query Match 13.0%; Score 34.8; DB 2; Length 272142;  
 Best Local Similarity 48.9%; Pred. No. 42;  
 Matches 93; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

QY 49 ATTACGAGTTGGCGTCAATCAATTAATGTAACCGTTCCGACCAAGAACTTCAGGAAG 108  
 Db 106761 ATTACCATGTTGAGTTTCATCTTATGTGATGACTTAATTCATAATTAATCTGAGAAA 106702  
 QY 109 TGGCGGGTTGCGTTCGGCGTCACCAATCGCGTCAAAAATCAGCAACCGCAGCAATCAAGG 168  
 Db 106701 TTCTATTGACTCTAGTTTCACACAGATGTGAGATAAAGAAATACCATATCTTTT 106642  
 QY 169 CATACCATCAACTTTTACCGCAACTGCGTAAGCGCGTGACCAATGCTATGTATCCAGT 228  
 Db 106641 TATTACATTCATTCTTCAAAAATTTAGGACGAGCTTGAGCAGTACTTGCTAACAAGA 106552  
 QY 229 GTAAATCAGCA 238  
 Db 106581 GTGATAAGCA 106572

RESULT 14  
 BD147361  
 LOCUS  
 DEFINITION  
 Primer for synthesizing full-length cDNA and use thereof.  
 ACCESSION  
 BD147361  
 VERSION  
 BD147361.1 GI:27853119  
 KEYWORDS  
 JP 2002191363-A/2204.  
 SOURCE  
 Homo sapiens (human)  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 Ota, T., Isogai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J.,  
 Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.  
 Primer for synthesizing full-length cDNA and use thereof  
 Patent: JP 2002191363-A 2204 09-JUL-2002;  
 JOURNAL  
 HELIX RESEARCH INSTITUTE  
 COMMENT  
 OS Homo sapiens (human)  
 PN JP 2002191363-A/2204  
 PD 09-JUL-2002  
 PF 28-JUL-2000 JP 2000280990  
 PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU  
 SUGIYAMA, T., WAKAMATSU, A., NAGAI, K. and OTSUKI, T.  
 PI JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU,  
 KEIICHI NAGAI, TETSUO OTSUKI  
 PC  
 C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ PC





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 14, 2003, 23:55:19 : Search time 172.322 Seconds  
(without alignments)  
4182.570 Million cell updates/sec

Title: US-09-928-457-89  
Perfect score: 267  
Sequence: 1 AATTATGACACACGATCA.....TTACCACTTCGATGCAATT 267

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq 19Jun03.\*  
1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*  
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*  
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*  
25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	267	100.0	267	19 AAV03597	Neisseria meningit
2	267	100.0	267	11 AAA15394	Genomic fragment o
3	238.2	89.2	46593	21 AAA81456	N. meningitidis pa
4	238.2	89.2	349980	21 AAF21612	Neisseria meningit
5	238.2	89.2	837096	21 AAA81489	N. meningitidis pa
6	235.8	88.3	339	21 AAA76029	Neisseria meningit
7	235.8	88.3	339	21 AAA15300	DNA encoding a pol
8	226.2	84.7	12077	21 AAA81734	N. meningitidis pa

ALIGNMENTS

RESULT 1  
AAV03597  
ID AAV03597 standard; DNA; 267 BP.

AC AAV03597;  
DE 22-OCT-1998 (first entry)  
DT Neisseria meningitidis DNA sequence E29.

DE N. gonorrhoeae; N. lactamica; chromosome 22491; region 1; region 2;  
region 3; pathogenicity; blood-brain barrier; diagnosis; infection;  
meningitis; ss.

OS Neisseria meningitidis.

XX WO9802547-A2.

PD 22-JAN-1998.

PF 11-JUL-1997; 97WO-FR01295.

PR 12-JUL-1996; 96FR-0008768.

XX (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.  
PA (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.  
PA (SMIK ) SMITHKLINE BEECHAM.

XX Nassif X, Tinsley C, Achtman M, Merker P, Ruelle J;  
PI Vinals C;

N. gonorrhoeae nuc  
N. gonorrhoeae nuc  
Human cDNA clone (i  
Human cDNA #506 di  
Glutamine:fructose  
Glutamine:fructose  
Human cDNA sequenc  
Glutamine:fructose  
Human GPAT II codi  
GPAT related nucle  
Nucleotide sequenc  
Human cDNA sequenc  
GPAT related nucle  
Human cDNA #895 di  
DNA encoding novel  
GPAT related nucle  
Listeria monocytog  
Listeria monocytog  
Oligonucleotide fo  
Bovine embryonic g  
Genomic sequence o  
Human ovarian can  
Listeria monocytog  
Drosophila melanog  
Drosophila melanog  
Listeria monocytog  
Human polynucleoti  
Human polynucleoti  
Listeria monocytog  
Transporters and i  
Transporters and i  
Human membrane tra  
Mouse ischaemic co  
DNA encoding novel  
DNA encoding novel  
Mouse GPAT II codi



```

XX
PR 09-OCT-1998; 98US-0103794.
PR 30-APR-1999; 99US-0132068.
XX
XX (CHIR ) CHIRON CORP.
XX
XX Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC,
PI Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;
PI Rappuoli R, Pizza M;
XX
XX WP1; 2000-318079/27.
XX
XX Isolated nucleotide sequences of Neisseria meningitidis which can be
PI used in the diagnosis and treatment of N. meningitidis infection and
PT other Neisserial infections, for example, N.gonorrhoea -
XX
XX Claim 7; Page 260-274; 1780pb; English.
XX
XX The present invention describes methods of obtaining immunogenic
CC proteins from Neisseria genomic sequences. AA81453 to AA82414
CC represent specifically claimed Neisseria meningitidis genomic DNA
CC sequences; AA81260 to AA81303 and AA825620 to AA825663 represent
CC Neisseria DNA sequences and their corresponding proteins; AA81254 to
CC AA81259 and AA81304 to AA81321 represent PCR primers used in the
CC isolation of Neisseria meningitidis DNA sequences; and AA81322 to
CC AA81452 represent Neisseria meningitidis MenB polynucleotide ORF
CC sequences, which are all used in the exemplification of the present
CC invention. The nucleic acid sequences, protein sequences, and antibodies
CC against them, can be used in the manufacture of a composition. The
CC composition can be used as a medicament (or in the manufacture of a
CC medicament) for treating, preventing or diagnosing infection due to
CC Neisserial bacteria. For example, some of the identified proteins could
CC be components of vaccines against Meningococcus B; against all serotypes;
CC and/or against all pathogenic Neissariae. Identification of sequences
CC from the bacterium will also facilitate production of biological probes,
CC particularly organism-specific probes. Attempts to make efficacious
CC Meningococcus B vaccines have failed mainly due to antigen tolerance.
CC Multivalent vaccines have also been tried but none have successfully
CC overcome antigenic variability. The provision of further, complete
CC sequences may provide an opportunity to identify secreted or surface
CC exposed proteins that may be presumed targets for the immune system and
CC which are not antigenically variable or at least more conserved than
CC other more variable regions.
XX
XX Sequence 46593 BP; 11355 A; 13195 C; 11355 G; 10687 T; 1 other;
SQ
Query Match 89.2%; Score 238.2; DB 21; Length 46593;
Best Local Similarity 97.4%; Pred. No. 1.3e-69;
Matches 264; Conservative 0; Mismatches 3; Indels 4; Gaps 2;
QY 1 AATTATGAACACACGCATCATCGTTTCGGCTCGTTTCGTTCGGTTCGCATTAGCAGGTTG 60
Db 46140 AATTATGAACACACGCATCATCGTTTCGGCTCGTTTCGTTCGGTTCGCATTAGCAGGTTG 46081
QY 61 CGGCTCAATCAATAATGTAAACCGTTTCCGACACAGAAACTTCAGGAACGTCGCCGTTTGC 120
Db 46080 CGGCTCAATCAATAATGTAAACCGTTTCCGACTAGAACTTCAGGAACGTCGCCGTTTGC 46021
QY 121 CTTGGGCGTCA---CCAATGCCGTAAATAATCAGCAACCGCAGCAATGAAGGCATACGCAT 177
Db 46020 CTTGGGCGTCAGCCCAATCGGTAAATCAGCAACCGCAGCAATGAAGGCATACGCAT 45961
QY 178 CAACCTTACCGCAACTGTGGTAAGCGCGTGA-CAATGCTATGTTACCAAGTGAATCAG 236
Db 45960 CAACCTTACCGCAACTGTGGTAAGCGCGTGAAGCAATGCTATGTACCAAGTGAATCAG 45901
QY 237 CACAATTCGGCGTTTACCACCTTCGATCAATT 267
Db 45900 CACAATTCGGCGTTTACCACCTTCGATCAATT 45870
XX
XX RESULT 4
XX AAF21612/c

```





XX PS Claim 7; Page 1565-1568; 1760pp; English.

XX CC The present invention describes methods of obtaining immunogenic

XX CC proteins from *Neisseria* genomic sequences. AAA81453 to AAA82414

XX CC represent specifically claimed *Neisseria meningitidis* genomic DNA

XX CC sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent

XX CC *Neisseria* DNA sequences and their corresponding proteins; AAA81254 to

XX CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the

XX CC isolation of *Neisseria meningitidis* DNA sequences; and AAA81322 to

XX CC AAA81452 represent *Neisseria meningitidis* MenB polynucleotide ORF

XX CC sequences, which are all used in the exemplification of the present

XX CC invention. The nucleic acid sequences, protein sequences, and antibodies

XX CC against them, can be used in the manufacture of a composition. The

XX CC composition can be used as a medicament (or in the manufacture of a

XX CC medicament) for treating, preventing or diagnosing infection due to

XX CC *Neisserial* bacteria. For example, some of the identified proteins could

XX CC be components of vaccines against *Meningococcus* B; against all serotypes;

XX CC and/or against all pathogenic *Neisseriae*. Identification of sequences

XX CC from the bacterium will also facilitate production of biological probes,

XX CC particularly organism-specific probes. Attempts to make efficacious

XX CC *Meningococcus* B vaccines have failed mainly due to antigen tolerance.

XX CC Multivalent vaccines have also been tried but none have successfully

XX CC overcome antigenic variability. The provision of further, complete

XX CC sequences may provide an opportunity to identify secreted or surface

XX CC exposed proteins that may be presumed targets for the immune system and

XX CC which are not antigenically variable or at least more conserved than

XX CC other more variable regions.

XX SQ Sequence 12077 BP; 2643 A; 3404 C; 3147 G; 2881 T; 2 other;

Query Match 84.7%; Score 226.2; DB 21; Length 12077;

Best Local Similarity 97.0%; Pred. No. 8.2e-66;

Matches 263; Conservative 0; Mismatches 3; Indels 5; Gaps 3;

QY 1 AATTATGAACACAGCATCATCGTTTCGGTGGCGTTCGTTGCGTTGCGATTACGAGTTG 60

DB 11627 AATTATGAACACAGCATCATCGTTTCGGTGGCGTTCGTTGCGTTGCGATTAGCA-GTTG 11685

QY 61 CGGCTCAATCAATATGTAACCGTTTCGACACAGAACTTCAGGAACGTCGCGGTTTCG 120

DB 11686 CGGCTCAATCAATATGTAACCGTTTCGACTAGAACTTCAGGAACGTCGCGGTTTCG 11745

QY 121 CTTGGGCGTCA---CGAATGCGGTAAATAATCAGCAACCGCAGCAATGAAGCATACGCAT 177

DB 11746 CTTGGGCGTCAAGCCCAATGCGGTAAATAATCAGCAACCGCAGCAATGAAGCATACGCAT 11805

QY 178 CAACCTTACCGCACTGTGGGTAGCGGTGA-CCAATGCTATGTTACCAAGTGAATCAG 236

DB 11806 CAACCTTACCGCACTGTGGGTAGCGGTGAAGCAATGCTATGTTACCAAGTGAATCAG 11865

QY 237 CACAATCGCGCTTACCACTTCGATGCAATT 267

DB 11866 CACAATCGCGCTTACCACTTCGATGCAATT 11896

RESULT 9

ABZ38335/c

ID ABZ38335 standard; DNA; 420 BP.

XX AC ABZ38335;

XX DT 07-MAR-2003 (first entry)

XX DE N. gonorrhoeae nucleotide sequence SEQ ID 1259.

XX KW Antibacterial; infection; vaccine; gene therapy; gene; ds.

XX OS *Neisseria gonorrhoeae*.

XX PN WO200279243-A2.

XX PD 10-OCT-2002.

XX PF 12-FEB-2002; 2002WO-IB02069.

XX PR 12-FEB-2001; 2001GB-0003424.

XX PA (CHIR-) CHIRON SPA.

XX PI Fontana MR, Pizza M, Masignani V, Monaci E;

XX WP1; 2003-058415/05.

XX P-PSDB; ABP77365.

DR New protein from *Neisseria gonorrhoeae*, useful for the manufacture of a

PT medicament for treating or preventing N. gonorrhoeae infection

XX PS Disclosure; Page 279; 815pp; English.

XX CC The present invention relates to proteins from *Neisseria gonorrhoeae*.

XX CC Also disclosed are the nucleic acid molecules encoding the proteins and

XX CC antibodies that specifically bind to the proteins. The composition

XX CC comprising the protein, nucleic acid or antibody is useful for the

XX CC manufacture of a medicament for treating or preventing N. gonorrhoeae

XX CC infection, this may be in the form of a vaccine or gene therapy.

XX CC Sequences given in records ABZ37706-ABZ42016 represent nucleic acid

XX CC molecules of the invention.

XX SQ Sequence 420 BP; 126 A; 78 C; 90 G; 126 T; 0 other;

Query Match 27.9%; Score 74.4; DB 25; Length 420;

Best Local Similarity 97.0%; Pred. No. 5.5e-15;

Matches 97; Conservative 0; Mismatches 1; Indels 2; Gaps 2;

QY 169 CATACGCATCAACTTACCGCAACTGTGGTAAGCGGTGA-CCAATGCTATGTTACGAG 227

DB 330 CACACGCATCAACTTACCGCAACTGTGGTAAGCGGTGAAGCCCAATGCTATGTTACGAG 272

QY 228 TGTATACGCAACATCGCGTTACCACTTCGATGCAATT 267

DB 271 TGTATACGCAACATCGCGTTACCACTTCGATGCAATT 232

RESULT 10

ABZ38334

ID ABZ38334 standard; DNA; 144 BP.

XX AC ABZ38334;

XX DT 07-MAR-2003 (first entry)

XX DE N. gonorrhoeae nucleotide sequence SEQ ID 1257.

XX KW Antibacterial; infection; vaccine; gene therapy; gene; ds.

XX OS *Neisseria gonorrhoeae*.

XX PN WO200279243-A2.

XX PD 10-OCT-2002.

XX PF 12-FEB-2002; 2002WO-IB02069.

XX PR 12-FEB-2001; 2001GB-0003424.

XX PA (CHIR-) CHIRON SPA.

XX PI Fontana MR, Pizza M, Masignani V, Monaci E;

XX WP1; 2003-058415/05.

XX P-PSDB; ABP77364.

XX PT New protein from *Neisseria gonorrhoeae*, useful for the manufacture of a

PT medicament for treating or preventing N. gonorrhoeae infection

XX

PS Disclosure; Page 279; 815pp; English.

CC The present invention relates to proteins from *Neisseria gonorrhoeae*.

CC Also disclosed are the nucleic acid molecules encoding the proteins and

CC antibodies that specifically bind to the proteins. The composition

CC comprising the protein, nucleic acid or antibody is useful for the

CC manufacture of a medicament for treating or preventing *N. gonorrhoeae*

CC infection. This may be in the form of a vaccine or gene therapy.

CC Sequences given in records AB237706-AB242016 represent nucleic acid

CC molecules of the invention.

XX

SQ Sequence 144 BP; 39 A; 34 C; 40 G; 31 T; 0 other;

Query Match 23.6%; Score 63; DB 25; Length 144;

Best Local Similarity 98.7%; Pred. No. 2.6e-11;

Matches 74; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 194 GTGGTAAGCGCGTGA-CAAATGCTATGTTACCAAGTGTATTCAGCAACAATCGGCGTTACC 252

Db 1 GTGGTAAGCGCGTGAAGCAATGCTATGTTACCAAGTGTATTCAGCAACAATCGGCGTTACC 60

QY 253 ACTTCCGATGCAATT 267

Db 61 ACTTCCGATGCAATT 75

RESULT 11

AAH05369

ID AAH05369 standard; cDNA; 827 BP.

AC AAH05369;

DT 26-JUN-2001 (first entry)

XX Human cDNA clone (5'-primer) SEQ ID NO:2204.

DE Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX Homo sapiens.

OS EP1074617-A2.

PN 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

PF 29-JUL-1999; 99JP-0248036.

PR 27-AUG-1999; 99JP-0300253.

PR 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

PA Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

DR Primer sets for synthesizing polynucleotides, particularly the 5602

PT full-length cDNAs defined in the specification, and for the detection

PT and/or diagnosis of the abnormality of the proteins encoded by the

PT full-length cDNAs -

XX Claim 1; SEQ ID 2204; 2537pp + CD ROM; English.

PS The present invention describes primer sets for synthesizing 5602

XX full-length cDNAs defined in the specification. Where a primer set

CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary

CC to the complementary strand of a polynucleotide which comprises one of

CC the 5602 nucleotide sequences defined in the specification, where the

CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination

CC of an oligonucleotide comprising a sequence complementary to the

CC complementary strand of a polynucleotide which comprises a 5'-end

CC sequence and an oligonucleotide comprising a sequence complementary to a

CC polynucleotide which comprises a 3'-end sequence, where the

CC oligonucleotide comprises at least 15 nucleotides and the combination of

CC the 5'-end sequence/3'-end sequence is selected from those defined in

CC the specification. The primer sets can be used in antisense therapy and

CC in gene therapy. The primers are useful for synthesising polynucleotides,

CC particularly full-length cDNAs. The primers are also useful for the

CC detection and/or diagnosis of the abnormality of the proteins encoded by

CC the full-length cDNAs. The primers allow obtaining of the full-length

CC cDNAs easily without any specialised methods. AAH0166 to AAH13628 and

CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to

CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632

CC represent oligonucleotides, all of which are used in the exemplification

CC of the present invention.

XX

SQ Sequence 827 BP; 200 A; 207 C; 234 G; 183 T; 3 other;

Query Match 13.0%; Score 34.6; DB 22; Length 827;

Best Local Similarity 55.4%; Pred. No. 0.23;

Matches 67; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 100 TCAGGAACGTGCGCGCTTTGCCTTGGCGGTCCACCAATGCGTAAATAACAGCAACCGCAG 159

Db 363 TAAGGACCGCGCGCTCTCACCGTGGCGTCCACCAACCGTGGCGAGTCCCATCTCTCG 422

QY 160 CAATGAAGGCATACGCATCACTTTACCGCACTGTGGTAAAGCGGTGACCAATGCTAT 219

Db 423 CGAGACCGACTGCGGCGTCCACATCAACGCGCGGAGATCGGCGTGGCCAGCACCAA 482

QY 220 G 220

Db 483 G 483

RESULT 12

ABX63506

ID ABX63506 standard; cDNA; 1651 BP.

AC ABX63506;

XX 26-FEB-2003 (first entry)

DT Human cDNA #506 differentially expressed in activated vascular tissue.

DE Human; gene; ss; vascular tissue; cytostatic; atherosclerosis;

XX cardiant; hypotensive; antidiabetic; gynaecological; vasotropic;

XX cerebroprotective; gene therapy; vascular disease; cancer; coronary;

XX artery disease; hypertension; diabetes; pre-eclampsia; restenosis;

XX ischaemia-reperfusion injury; stroke;

XX Homo sapiens.

OS US2002137081-A1.

PN 26-SEP-2002.

XX 08-JAN-2002; 2002US-0044090.

XX 28-JUL-2000; 2000US-222469P.

PR 08-JAN-2001; 2001US-260483P.

XX (BAND/) BANDMAN O.

PA Bandman O;

XX WPI; 2003-110597/10.

DR Combination for diagnosing, staging, treating, or monitoring the

PT progression of treatment of a vascular disease, e.g. atherosclerosis,

PT comprises several cDNAs that are differentially expressed in activated

PT vascular tissue -

Claim 1; Page -; 18pp; English.

PS This invention relates to a combination comprising several cDNAs that  
 XX are differentially expressed in activated vascular tissue. The invention  
 CC also discloses a high throughput method for detecting differentially  
 CC expressed cDNAs in a sample. The cDNAs of the invention may have  
 CC antiatherosclerotic; cytostatic; cardiant; hypotensive; antidiabetic;  
 CC gynaecological; vasotropic and cerebroprotective activities and may be  
 CC used in gene therapy. The cDNAs of the invention may be used in a  
 CC high-throughput methods for detecting differential expression of one or  
 CC more cDNAs in a sample, or screening several molecules or compounds to  
 CC identify a molecule or compound that specifically binds a cDNA of the  
 CC invention. A protein encoded by the cDNA may be used to screen several  
 CC molecules or compounds to identify a ligand that specifically binds to  
 CC the protein, or to produce or purify an antibody to the protein that can  
 CC be used to detect a protein in a sample or purify a natural or  
 CC recombinant protein from a sample. The nucleotides may be useful for  
 CC diagnosing, staging, treating, or monitoring the progression of  
 CC treatment of a vascular disease, e.g. atherosclerosis, cancer, coronary  
 CC artery disease, hypertension, diabetes, pre-eclampsia, ischaemia-  
 CC reperfusion injury, restenosis, or stroke. The cDNAs can also be used  
 CC for large-scale genetic or gene expression analysis of several new  
 CC nucleic acid molecules. Antibodies to the proteins encoded by the  
 CC cDNAs are useful for diagnosing pre-pathologic disorders, and chronic  
 CC or acute diseases associated with abnormalities in the expression,  
 CC amount or distribution of the protein. The present sequence  
 CC represents a cDNA of the invention that is differentially expressed in  
 CC activated vascular tissue.

CC Note: The sequence data for this patent did not form part of the  
 CC specification, but was obtained in electronic format directly from USPTO  
 CC at <http://seqdata.uspto.gov/sequence.html?DocID=20020137081>.

XX Sequence 1651 BP; 386 A; 441 C; 410 G; 414 T; 0 other;

Query Match 13.0%; Score 34.6; DB 25; Length 1651;  
 Best Local Similarity 55.4%; Pred. No. 0.3;  
 Matches 67; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 100 TCAGGACGTCGCGGTTGCGTTGGCGTCACCAATGCCGTAAATACAGCAACCGCAG 159  
 Db 207 TAAGGACCGCGCGCTCTCACCGTGGCGGTCCACCAACACCGTGGCGAGCTCTCTCG 266.

QY 160 CAATCAAGGATAGCGATCACTTTACCGCAACTGTGGGTAAAGCGGTGACCAATGCTAT 219  
 Db 267 CGAGACCGACTGCGGGTCCACATCAACGAGGCGCGGAGATCGGCGTGGCCAGCACCA 326

QY 220 G 220  
 Db 327 G 327

RESULT 13  
 AAV18994  
 ID AAV18994 standard; cDNA; 1796 BP.

AC AAV18994;

DT 17-AUG-1998 (first entry)

XX Glutamine:fructose-6-phosphate amidotransferase TGC028-3 probe.

XX Glutamine:fructose-6-phosphate amidotransferase; GFAT; TGC028-3;  
 KW human; hypoglycaemia; diabetes; antidiabetic; therapy; probe; ss.

XX Homo sapiens.

OS EP824149-A2.

PN 18-FEB-1998.

XX 13-AUG-1997; 97EP-0113934.

XX 13-AUG-1996; 96JP-0213944.

XX (TAKE ) TAKEDA CHEM IND LTD.

XX Hikichi Y, Nishi K, Shintani Y;

XX WPI; 1998-122309/12.

XX Glutamine:fructose-6-phosphate amidotransferase proteins - useful  
 PT for treating hypoglycaemia or to screen for antidiabetic agents

XX Example 3; Page 51-52; 77pp; English.

XX This sequence comprises a partial sequence of TGC028-3 (see AAV18998)  
 CC cDNA and codes for a portion of a novel human protein (see AAW3771)  
 CC having glutamine:fructose-6-phosphate amidotransferase (GFAT)  
 CC activity. It has been labelled with (alpha-32P)dCTP and used as a  
 CC probe to determine tissue specificity of gene expression by  
 CC Northern blotting. The invention relates to novel proteins (see  
 CC AAW29526, AAW3771-72) having GFAT activity, and DNAs encoding them (see  
 CC AAV18997-89), and their use for treating hypoglycaemia and for  
 CC identifying antidiabetic agents.

XX Sequence 1796 BP; 476 A; 451 C; 493 G; 376 T; 0 other;

Query Match 13.0%; Score 34.6; DB 19; Length 1796;  
 Best Local Similarity 55.4%; Pred. No. 0.31;  
 Matches 67; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 100 TCAGGACGTCGCGGTTGCGTTGGCGGTCAACCAATGCCGTAAATACAGCAACCGCAG 159  
 Db 1319 TAAGGACCGCGCGCTCTCACCGTGGCGGTCCACCAACACCGTGGCGAGCTCTCTCG 1378

QY 160 CAATCAAGGATAGCGATCACTTTACCGCAACTGTGGGTAAAGCGGTGACCAATGCTAT 219  
 Db 1379 CGAGACCGACTGCGGGTCCACATCAACGAGGCGCGGAGATCGGCGTGGCCAGCACCA 1438

QY 220 G 220

Db 1439 G 1439

RESULT 14

AAV18988

ID AAV18988 standard; cDNA; 1845 BP.

AC AAV18988;

DT 17-AUG-1998 (first entry)

XX Glutamine:fructose-6-phosphate amidotransferase TGC028-3 cDNA.

XX Glutamine:fructose-6-phosphate amidotransferase; GFAT; TGC028-3;  
 KW human; hypoglycaemia; diabetes; antidiabetic; therapy; ds.

XX Homo sapiens.

XX EP824149-A2.

PN 18-FEB-1998.

XX 13-AUG-1997; 97EP-0113934.

XX 13-AUG-1996; 96JP-0213944.

XX (TAKE ) TAKEDA CHEM IND LTD.

XX Hikichi Y, Nishi K, Shintani Y;

XX WPI; 1998-122309/12.

XX P-PSDB; AAW37711.

XX Glutamine:fructose-6-phosphate amidotransferase proteins - useful  
 PT for treating hypoglycaemia or to screen for antidiabetic agents



XX PS Claim 7; Page 43-44; 77pp; English.

XX CC This cDNA sequence codes for TGC028-3 (see AAW37771), a novel human

XX CC protein having glutamine:fructose-6-phosphate amidotransferase

XX CC (GFAT) activity and showing 74% homology to human GFAT. To isolate

XX CC TGC028-3 cDNA, a human brain cDNA library was probed with an

XX CC oligonucleotide (AAV18993), and DNA from a positive clone was

XX CC amplified using the same oligonucleotide as primer and introduced

XX CC into *Escherichia coli* DH10B. Transformants were screened by colony

XX CC PCR (see also AAV18991-92), and a colony which formed a 512-bp

XX CC amplified fragment was selected. The insert DNA of a plasmid from

XX CC this positive strain included the 1845 open reading frame of

XX CC TGC028-3. Plasmid pTBE1943 containing this DNA has been introduced

XX CC into *E. coli* DH10B and deposited as FERM BP-5622 (IPO 16004).

XX CC Also claimed are: a recombinant vector containing DNA (see also

XX CC AAV18987 and AAV18989) coding for a protein having GFAT activity (see

XX CC also AAW29526 and AAW37772) or its partial peptide, a transformant

XX CC transformed with the vector, a method for producing the protein by

XX CC culturing the transformant, an antibody against the protein, a

XX CC screening assay for compounds that inhibit the GFAT activity of the

XX CC protein, and a compound that inhibits GFAT activity. The GFAT, the

XX CC peptide fragment, or DNA coding for them, can be used in the

XX CC treatment and prophylaxis of hypoglycaemia. The GFAT inhibitor can

XX CC be used to treat diabetes.

XX SQ Sequence 1845 BP; 486 A; 459 C; 505 G; 385 T; 0 other;

Query Match 13.0%; Score 34.6; DB 19; Length 1845;

Best Local Similarity 55.4%; Pred. No. 0.32;

Matches 67; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 100 TCAGGACGTGCGCGTTTGCTTGGCGGTCCACCAATGCCGTAATAAATCAGCAACCGCAG 159

Db 1311 TAAGGACCGCGCGCTCTCACCGTGGCGGTCCACCAACCGTGGCGCTCTCTCG 1370

QY 160 CAATGAAGGCATACGCATCACTTTACCGCACTGTGGTAAGCGGTGACCAATGCTAT 219

Db 1371 CGAGACCGACTGCGCGGTCCACATCAACGCGGCGGAGATCGCGTGGCCAGCACCA 1430

QY 220 G 220

Db 1431 G 1431

RESULT 15

AAH14073

ID AAH14073 standard; cDNA; 1966 BP.

AC AAH14073;

XX 26-JUN-2001 (first entry)

XX Human cDNA sequence SEQ ID NO:11221.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.

XX 27-AUG-1999; 99JP-0300253.

XX 11-JAN-2000; 2000JP-0118776.

XX 02-MAY-2000; 2000JP-0183767.

XX 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX FA

XX XX

PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S; Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX PR Primer sets for synthesizing polynucleotides, particularly the 5602

XX PT full-length cDNAs defined in the specification, and for the detection

XX PT and/or diagnosis of the abnormality of the proteins encoded by the

XX PT full-length cDNAs -

XX Claim 8; SEQ ID 11221; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesising 5602

XX full-length cDNAs defined in the specification. Where a primer set

XX comprises: (a) an oligo-dr primer and an oligonucleotide complementary

XX to the complementary strand of a polynucleotide which comprises one of

XX the 5602 nucleotide sequences defined in the specification, where the

XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination

XX of an oligonucleotide comprising a sequence complementary to the

XX complementary strand of a polynucleotide which comprises a 5'-end

XX sequence and an oligonucleotide comprising a sequence complementary to a

XX polynucleotide which comprises a 3'-end sequence, where the

XX oligonucleotide comprises at least 15 nucleotides and the combination of

XX the 5'-end sequence/3'-end sequence is selected from those defined in

XX the specification. The primer sets can be used in antisense therapy and

XX in gene therapy. The primers are useful for synthesising polynucleotides,

XX particularly full-length cDNAs. The primers are also useful for the

XX detection and/or diagnosis of the abnormality of the proteins encoded by

XX the full-length cDNAs. The primers allow obtaining of the full-length

XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and

XX AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to

XX AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632

XX represent oligonucleotides, all of which are used in the exemplification

XX of the present invention.

XX SQ Sequence 1966 BP; 471 A; 516 C; 472 G; 507 T; 0 other;

Query Match 13.0%; Score 34.6; DB 22; Length 1966;

Best Local Similarity 55.4%; Pred. No. 0.33;

Matches 67; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 100 TCAGGACGTGCGCGTTTGCTTGGCGGTCCACCAATGCCGTAATAAATCAGCAACCGCAG 159

Db 363 TAAGGACCGCGCGCTCTCACCGTGGCGGTCCACCAACCGTGGCGGTCTCTCG 422

QY 160 CAATGAAGGCATACGCATCACTTTACCGCACTGTGGTAAGCGGTGACCAATGCTAT 219

Db 423 CGAGACCGACTGCGCGGTCCACATCAACGCGGCGGAGATCGCGTGGCCAGCACCA 482

QY 220 G 220

Db 483 G 483

Search completed: November 15, 2003, 00:35:42

Job time : 177.322 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 14, 2003, 23:58:39 ; Search time 1781.18 Seconds  
(without alignments)  
3643.257 Million cell updates/sec

Title: US-09-928-457-89

Perfect score: 267

Sequence: 1 AATTATGACACACCATCA.....TTACACTTCGATGCAATT 267

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152338056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hci:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rtd:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	37.4	14.0	687	10	BG492320
2	35.2	13.2	580	13	BQ852510
3	34.6	13.0	350	9	AV659367
4	34.6	13.0	585	10	BG333201

RESULT 1  
BG492320  
LOCUS 602536235F1 NIH\_MGC\_59 Homo sapiens CDNA clone IMAGE:4655216 5', linear  
DEFINITION mRNA sequence.  
ACCESSION BG492320  
VERSION BG492320.1 GI:13453832  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 687)  
NIH-MGC <http://mgc.nci.nih.gov/>  
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
TITLE Unpublished  
JOURNAL  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgabs-r@mail.nih.gov](mailto:cgabs-r@mail.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library Preparation: CLONTECH Laboratories, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.lnl.gov>  
Plate: LNCMI44 row: e column: 09  
High quality sequence stop: 548.

#### ALIGNMENTS

5	34.6	13.0	706	10	BG492320
6	34.6	13.0	827	9	AU124329
7	34.6	13.0	842	29	BZ729792
8	34.6	13.0	857	9	AU125108
9	34.6	13.0	911	13	BQ851308
10	33.8	12.7	811	28	BH422048
11	33.4	12.5	345	28	BH695730
12	33.4	12.5	661	12	B1165435
13	33.4	12.5	730	28	BH605016
14	33.4	12.5	883	10	BH313432
15	33.4	12.5	962	29	CNS02GGP
16	33	12.4	580	10	BF112720
17	33	12.4	939	13	EX430863
18	33	12.4	1041	12	BM423748
19	33	12.4	1201	13	EX385842
20	32.8	12.3	562	12	BI208441
21	32.8	12.3	624	12	BI930537
22	32.6	12.2	297	9	AV087447
23	32.6	12.2	443	12	BM030975
24	32.6	12.2	519	9	AW621988
25	32.6	12.2	1076	14	CB192570
26	32.4	12.1	733	28	BH140229
27	32.2	12.1	463	29	CNS013P2
28	32.2	12.1	746	12	B1154092
29	32.2	12.1	862	29	BZ787001
30	32.2	12.1	1201	13	EX324674
31	32	12.0	530	9	AW599664
32	32	12.0	803	28	BZ029668
33	32	12.0	804	28	BH364559
34	32	12.0	847	29	BZ961574
35	32	12.0	999	29	BZ493336
36	32	12.0	1021	29	CNS078BU
37	31.8	11.9	524	29	AG249081
38	31.8	11.9	589	12	B1167011
39	31.8	11.9	571	9	AA522229
40	31.8	11.9	681	28	BH527710
41	31.8	11.9	771	10	BF247178
42	31.8	11.9	810	28	BH481443
43	31.8	11.9	982	29	CNS0338Y
44	31.8	11.9	1201	13	EX438879
45	31.6	11.8	231	14	R29865

BG492320 602518007  
AU124329 AU124329  
BZ729792 OGFAT77TC  
AU125108 AU125108  
BQ851308 AGENCOURT  
BH422048 BOGLS36TF  
BH695730 BOMPE14TR  
B1165435 RE05337.5  
BH605016 BOH0A02TF  
BE313432 601148895  
AL156306 Tetraodon  
BF112720 EST440310  
EX430863 EX430863  
BM423748 AGENCOURT  
EX385842 EX385842  
BI208441 EST526481  
BI930537 EST550426  
AV087447 AV087447  
BM030975 A95552.WA  
AW621988 EST312786  
CB192570 AGENCOURT  
BH140229 ZMMBB000  
AL105728 Drosophila  
B1154092 602870993  
BZ787001 PUFHD79TB  
EX324674 EX324674  
AW599664 G951B06.Y  
BZ029668 oef79e06.  
BH364559 BOHIQ56TR  
BZ961574 PUDD82TD  
BZ493336 BONKX75TR  
AL433732 T3 end of  
AG249081 Lotus japonicus  
B1167011 RE07278.5  
AA522229 V143C09.1  
BH527710 BOGGL33TR  
BF247178 601857849  
BH481443 BOGRG76TF  
AL225835 Tetraodon  
EX438879 EX438879  
R29865 12470 Lambd

```

FEATURES
  source
    Location/Qualifiers
      1..687
        /organism="Homo sapiens"
        /mol_type="mRNA"
        /db_xref="taxon:9606"
        /clone="IMAGE:4655216"
        /tissue_type="mucoepidermoid carcinoma"
        /lab_host="DH10B (T1 phage-resistant)"
        /clone_lib="NHG_MGC_59"
        /note="Organ: lung; Vector: pDNR-LIB (Clontech); Site_1: SfiI (ggccattggcgc); Site_2: SfiI (ggccattggcgc); Double-stranded cDNA was prepared from cell line RNA. 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGCCATATGGCC-3' and 3' adaptor (where B = A, C, G and N = A, C, G, or T). Average insert size 1.65 kb (range 0.9-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MGC Library."
      BASE COUNT      197 a      166 c      146 g      178 t
      ORIGIN
        Query Match      14.0%; Score 37.4; DB 10; Length 687;
        Best Local Similarity 67.1%; Pred No. 0.52; Indels 0; Gaps 0;
        Matches 53; Conservative 0; Mismatches 26;

      QY      64      CTCATCAATATGTAACGGTTTCGACAGAACTTCAGGAAGTGGCGGTTTGCTT 123
      Db      561      CTCATTTATGATCTAATTTAAACGAGGATAAATTTCAAAACAGTGTCTTCCTTCCCTT 620

      QY      124      GCGGTCACCATGCGGTA 142
      Db      621      GGGCATCATCATTTCCATA 639

      RESULT 2
      LOCUS      BQ852510
      DEFINITION      BQ852510.y9.abl QG-ABCDI lettuce salinas Lactuca sativa CDNA clone
      ACCESSION      BQ852510
      VERSION      BQ852510.1 GI:22237975
      KEYWORDS      EST.
      SOURCE      Lactuca sativa
      ORGANISM      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
        Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
        asterids; campanulids; Asterales; Asteraceae; Cichorioideae;
        Cichorieae; Lactuca.
      REFERENCE      1 (bases 1 to 580)
      AUTHORS      Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Rieseberg, L.,
        Lin, H., van Damme, M., Lavelle, D., Chevalier, P., Ziegler, J., Ellison,
        P., Kolkmann, J., Slabaugh, K.S., Livingston, K., Zhou, Y., Lai, Z.,
        Church, S., Jackson, L. and Bradford, K.
      TITLE      Lettuce and Sunflower ESTs from the Compositae Genome Project
        http://comgenomics.ucdavis.edu/
      JOURNAL      Unpublished
      COMMENT      Contact: Alexander Kozik [R.W.Michelmore]
        Department of Vegetable Crops, R.W.Michelmore Lab
        University of California at Davis (UCD)
        Asmunsund Hall, UCD, Davis, CA 95616, USA
        Tel: 1-(530)-742-1742
        Fax: 1-(530)-752-9659
        Email: akozik@ucdavis.org [michelmore@vegmail.ucdavis.edu]
        singleton, see http://cgdb.ucdavis.edu/ for details.
        Plate: QGB18 row D column: 15.

      FEATURES
        source
          Location/Qualifiers
            1..580
              /organism="Lactuca sativa"
              /mol_type="mRNA"
              /cultivar="Salinas"

  /db_xref="taxon:4236"
  /clone="QGB18D15"
  /lab_host="E.coli"
  /clone_lib="QG-ABCDI lettuce salinas"
  /note="Vector: pBRCNDASFIAB; The library was constructed from 10 different sources of RNA from a single genotype. Separate cDNAs were generated using primers that incorporated unique 5' and 3' tags to distinguish each source of RNA. cDNAs were then pooled, size-fractionated, directionally cloned into a custom medium-copy vector and transformations made with four size classes to minimize size bias. Details of each source of RNA and library construction can be obtained at http://cgdb.ucdavis.edu/ TAG_LIB=QG-ABCDI lettuce salinas TAG_TISSUE=chemical induction TAG_SEQ=rgttagccggg"
  BASE COUNT      183 a      102 c      143 g      152 t
  ORIGIN
    Query Match      13.2%; Score 35.2; DB 13; Length 580;
    Best Local Similarity 58.7%; Pred No. 2.4;
    Matches 61; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

    QY      164      GAAGCATACGCATCAACTTTACCGCAACTGTGGTAAAGCGGTGACCAATGCTATGTTA 223
    Db      244      GAAACATCGCGATGATCTTGGTACCAATGTGGAGTTGGACTGCTACATGGAAGGAA 303

    QY      224      CCAGTGTATCAGCAATCGGGTTACCACTTCGATCGCAATT 267
    Db      304      CCAGTGGAGAATCAATCGGAATCAATTCGTGATCGCGTT 347

  RESULT 3
  LOCUS      AV659367
  DEFINITION      AV659367 GLC Homo sapiens cDNA clone GLCFWD02 3', mRNA sequence.
  ACCESSION      AV659367
  VERSION      AV659367.1 GI:9880381
  KEYWORDS      EST.
  SOURCE      Homo sapiens (human)
  ORGANISM      Homo sapiens
    Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  REFERENCE      1 (bases 1 to 350)
  AUTHORS      Xu, X., Huang, J., Xu, Z., Qian, B., Zhu, Z., Yan, Q., Cai, T., Zhang, X.,
    Xiao, H., Qu, J., Liu, F., Huang, Q., Cheng, Z., Li, N., Du, J., Hu, W.,
    Shen, K., Lu, G., Fu, G., Zhong, M., Xu, S., Gu, W., Huang, W., Zhao, X.,
    Hu, G., Gu, J., Chen, Z. and Han, Z.
  TITLE      Insight into hepatocellular carcinogenesis at transcriptome level
    by comparing gene expression profiles of hepatocellular carcinoma
    with those of corresponding noncancerous liver
  JOURNAL      Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
  MEDLINE      21625106
  PUBMED      11752456
  COMMENT      Contact: Zequang Han
    Chinese National Human Genome Center at Shanghai
    351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
    201203, P. R. China
    Tel: 86-21-50801919 (ex.45)
    Fax: 86-21-50801922
    Email: hanzg@chgc.sh.cn
    This clone is available at CHGC in Shanghai.

  FEATURES
    source
      Location/Qualifiers
        1..350
          /organism="Homo sapiens"
          /mol_type="mRNA"
          /db_xref="taxon:9606"
          /clone="GLCFWD02"
          /tissue_type="corresponding non cancerous liver tissue"
          /dev_stage="Adult"
          /lab_host="SCLR"
          /clone_lib="GLC"
          /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:

```

```

BASE COUNT      66 a      105 c      115 g      64 t
ORIGIN
Query Match      13.0%; Score 34.6; DB 9; Length 350;
Best Local Similarity 55.4%; Pred. No. 3;
Matches 67; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 100 TCAGGAACGTGCGCGCTTTGCTTGGCGTCCACCAATGCCGTAAATAATCAGCAACCGCAG 159
    |||||
DB 89 TAAGGACCGCGCGCTCTCACCGTGGCGTCCACCAACACCGTGGCGAGTCCATCTCTCG 148
    |||||

QY 160 CAATGAAGGCATACGCATCAACTTTACCGCAACTGTGGTAAAGCGCGTGACCAATGCTAT 219
    |||||
DB 149 CGAGACCGACTGCGCGGTCCACATCAACGAGCGCGGAGATCGCGGTGCCAGCACCAA 208
    |||||

QY 220 G 220
DB 209 G 209

RESULT 4
BG333201      585 bp      mRNA      linear      EST 27-FEB-2001
LOCUS      502430915F1 NIH_MGC_18 Homo sapiens cDNA clone IMAGE:4548674 5',
DEFINITION      mRNA sequence.
ACCESSION      BG333201
VERSION      BG333201.1 GI:13139639
KEYWORDS      EST.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 585)
AUTHORS      NIH-MGC http://mgc.nci.nih.gov/.
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished
COMMENT      Contact: Robert Strausberg, Ph.D.
            Email: cgabbs@mail.nih.gov
            Tissue Procurement: DCTD/DTF/Gazdar
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCMI238 row: n column: 03
            High quality sequence stop: 585.
            Location/Qualifiers
                1..585
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:4548674"
                /tissue_type="large cell carcinoma"
                /lab_host="DH10B (phage-resistant)"
                /clone_lib="NIH MGC 18"
                /note="Organ: lung; Vector: pOTB7; Site:1: XhoI; Site:2:
            EcoRI; cDNA made by oligo-dr priming. Directionally cloned
            into EcoRI/XhoI sites using the following 5' adaptor:
            GGCACGAG(G). Library constructed by Ling Hong in the
            laboratory of Gerald M. Rubin (University of California,
            Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
            Superscript II RT (Life Technologies). Note: this is a
            NIH_MGC Library."
BASE COUNT      138 a      148 c      165 g      134 t
ORIGIN
Query Match      13.0%; Score 34.6; DB 10; Length 585;
Best Local Similarity 55.4%; Pred. No. 3.8;
Matches 67; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 100 TCAGGAACGTGCGCGCTTTGCTTGGCGTCCACCAATGCCGTAAATAATCAGCAACCGCAG 159
    |||||

```

```

DB 375 TAAGGACCGCGCGCTCTCACCGTGGCGTCCACCAACACCGTGGCGAGTCCATCTCTCG 434
    |||||
QY 160 CAATGAAGGCATACGCATCAACTTTACCGCAACTGTGGTAAAGCGCGTGACCAATGCTAT 219
    |||||
DB 435 CGAGACCGACTGCGCGGTCCACATCAACGAGCGCGGAGATCGCGGTGCCAGCACCAA 494
    |||||

QY 220 G 220
DB 495 G 495

RESULT 5
BG489592      706 bp      mRNA      linear      EST 27-MAR-2001
LOCUS      602518007F1 NIH_MGC_18 Homo sapiens cDNA clone IMAGE:4636608 5',
DEFINITION      mRNA sequence.
ACCESSION      BG489592
VERSION      BG489592.1 GI:13451102
KEYWORDS      EST.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 706)
AUTHORS      NIH-MGC http://mgc.nci.nih.gov/.
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished
COMMENT      Contact: Robert Strausberg, Ph.D.
            Email: cgabbs@mail.nih.gov
            Tissue Procurement: DCTD/DTF/Gazdar
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCMI395 row: n column: 01
            High quality sequence stop: 650.
            Location/Qualifiers
                1..706
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:4636608"
                /tissue_type="large cell carcinoma"
                /lab_host="DH10B (phage-resistant)"
                /clone_lib="NIH MGC 18"
                /note="Organ: lung; Vector: pOTB7; Site:1: XhoI; Site:2:
            EcoRI; cDNA made by oligo-dr priming. Directionally cloned
            into EcoRI/XhoI sites using the following 5' adaptor:
            GGCACGAG(G). Library constructed by Ling Hong in the
            laboratory of Gerald M. Rubin (University of California,
            Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
            Superscript II RT (Life Technologies). Note: this is a
            NIH_MGC Library."
BASE COUNT      167 a      184 c      204 g      151 t
ORIGIN
Query Match      13.0%; Score 34.6; DB 10; Length 706;
Best Local Similarity 55.4%; Pred. No. 4.2;
Matches 67; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 100 TCAGGAACGTGCGCGCTTTGCTTGGCGTCCACCAATGCCGTAAATAATCAGCAACCGCAG 159
    |||||
DB 222 TAAGGACCGCGCGCTCTCACCGTGGCGTCCACCAACACCGTGGCGAGTCCATCTCTCG 281
    |||||

QY 160 CAATGAAGGCATACGCATCAACTTTACCGCAACTGTGGTAAAGCGCGTGACCAATGCTAT 219
    |||||
DB 282 CGAGACCGACTGCGCGGTCCACATCAACGAGCGCGGAGATCGCGGTGCCAGCACCAA 341
    |||||

QY 220 G 220

```

```

SOURCE          Zea mays
ORGANISM        Zea mays
REFERENCE       Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AUTHORS         Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
                clade; Panicoideae; Andropogoneae; Zea.
TITLE           1 (bases 1 to 842)
JOURNAL         Whitelaw C.A., Quackenbush J., Van Aken S., Utterback T., Resnick
COMMENT         R.W., Nunberg A., Budiman M.A., Becell J.A., Ronliffing T., Citek
                Consortium for Maize Genomics
                Other GSSs: OGFAT77TM
                Unpublished
                Contact: Cathy Whitelaw

FEATURES       source
               location/Qualifiers
               1..842
               /organism="Zea mays"
               /mol_type="genomic DNA"
               /strain="B73"
               /db_xref="taxon:4577"
               /clone="ZMMWma0241N10"
               /clone_lib="ZM 0.7-1.5 KB"
               /note="Vector: pCCK-S; Site I: HincII; 0.7-1.5 kb
               methylation filtered genomic DNA library"

BASE COUNT     191 a 193 c 213 g 245 t
ORIGIN
Query Match    13.0%; Score 34.6; DB 29; Length 842;
Best Local Similarity 49.7%; Pred.No. 4.5; Indels 0; Gaps 0;
Matches 88; Conservative 0; Mismatches 89;

QY 18 TCATCGTTTCGGCTGCGTTCGTTCGTTGCATTAGCAGTTCCGGCTCAATCAATAATG 77
DB 74 TCAACCCCTTTCCTATCTCGGACACTTAGTCCGGATGAGCCTTCTCTTAATCAATCGGG 133
QY 78 TAACGGITTCGACAGAACTTCAGAACGTGCCGGTTTGCTTGGGGGTCCACCAATG 137
DB 134 TCATTAAGACCCCGCAAGGACCCACACACTTAGGTGTTTCTTGTAGCTTTTACAAGC 193
QY 138 CGTAAAAAATCAGCAACCGCAGCAATGAAGGCATACGCATCAACTTTTACCGCAACTG 194
DB 194 ACTTAAGATAGAATGGGAAGGAGAGAGGCAATCCAGCAACAATAGCACAAATG 250

RESULT 8
LOCUS          AU125108 NT2RM4 Homo sapiens cDNA clone NT2RM4001049 5', mRNA
DEFINITION     sequence.
VERSION        AU125108.1 GI:10949824
KEYWORDS       EST.
SOURCE         Homo sapiens (human)
ORGANISM       Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE      1 (bases 1 to 857)
AUTHORS        Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Saito,K., Yamamoto,J.,
                Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S. and
                Isogai,T.
TITLE          HRI human cDNA project (Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S.,
                Saito,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki
                Y., Sugano,S., Isogai,T.)
JOURNAL        Unpublished
COMMENT        Contact: Takao Isogai
                Genomics Laboratory
                Helix Research Institute

SOURCE          Zea mays
ORGANISM        Zea mays
REFERENCE       Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AUTHORS         Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
                clade; Panicoideae; Andropogoneae; Zea.
TITLE           1 (bases 1 to 842)
JOURNAL         Whitelaw C.A., Quackenbush J., Van Aken S., Utterback T., Resnick
COMMENT         R.W., Nunberg A., Budiman M.A., Becell J.A., Ronliffing T., Citek
                Consortium for Maize Genomics
                Other GSSs: OGFAT77TM
                Unpublished
                Contact: Cathy Whitelaw

FEATURES       source
               location/Qualifiers
               1..842
               /organism="Zea mays"
               /mol_type="genomic DNA"
               /strain="B73"
               /db_xref="taxon:4577"
               /clone="ZMMWma0241N10"
               /clone_lib="ZM 0.7-1.5 KB"
               /note="Vector: pCCK-S; Site I: HincII; 0.7-1.5 kb
               methylation filtered genomic DNA library"

BASE COUNT     191 a 193 c 213 g 245 t
ORIGIN
Query Match    13.0%; Score 34.6; DB 29; Length 842;
Best Local Similarity 49.7%; Pred.No. 4.5; Indels 0; Gaps 0;
Matches 88; Conservative 0; Mismatches 89;

QY 18 TCATCGTTTCGGCTGCGTTCGTTCGTTGCATTAGCAGTTCCGGCTCAATCAATAATG 77
DB 74 TCAACCCCTTTCCTATCTCGGACACTTAGTCCGGATGAGCCTTCTCTTAATCAATCGGG 133
QY 78 TAACGGITTCGACAGAACTTCAGAACGTGCCGGTTTGCTTGGGGGTCCACCAATG 137
DB 134 TCATTAAGACCCCGCAAGGACCCACACACTTAGGTGTTTCTTGTAGCTTTTACAAGC 193
QY 138 CGTAAAAAATCAGCAACCGCAGCAATGAAGGCATACGCATCAACTTTTACCGCAACTG 194
DB 194 ACTTAAGATAGAATGGGAAGGAGAGAGGCAATCCAGCAACAATAGCACAAATG 250

RESULT 8
LOCUS          AU125108 NT2RM4 Homo sapiens cDNA clone NT2RM4001049 5', mRNA
DEFINITION     sequence.
VERSION        AU125108.1 GI:10949824
KEYWORDS       EST.
SOURCE         Homo sapiens (human)
ORGANISM       Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE      1 (bases 1 to 857)
AUTHORS        Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Saito,K., Yamamoto,J.,
                Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S. and
                Isogai,T.
TITLE          HRI human cDNA project (Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S.,
                Saito,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki
                Y., Sugano,S., Isogai,T.)
JOURNAL        Unpublished
COMMENT        Contact: Takao Isogai
                Genomics Laboratory
                Helix Research Institute

```

1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
Tel: 81-438-52-3975  
Fax: 81-438-52-3986  
Email: genomics@rii.co.jp

HRI human cDNA project; 5'- and 3'-end one pass sequencing: Helix Research Institute; cDNA library construction: Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute.

#### FEATURES

Location/Qualifiers

1..857

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="NT2RM4001049"

/cell\_type="teratocarcinoma"

/cell\_line="NT2"

/clone\_lib="NT2RM4"

/note="Vector: pME189FL3; mRNA from uninduced NT2 neuronal precursor cells"

201 a 217 c 238 g 196 t 5 others

#### BASE COUNT

ORIGIN

Query Match 13.0%; Score 34.6; DB 9; Length 857;

Best Local Similarity 55.4%; Pred. No. 4.5; Indels 0; Gaps 0;

Matches 67; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 100 TCAGGAACGTCGCCGCTTGCCTTGGCGGTCCACCAATGCCGTAAATAATCAGCAACCGCAG 159

Db 349 TAAGGACCGCGCGCTCTCACCGTGGCGTCACCAACCGTGGCGAGCTCCATCTCTCG 408

QY 160 CAATGAAGGATAGCATCACTTACCGCACTGTGGTAAAGCGGTGACCAATGCTAT 219

Db 409 CGAGACCGACTGCGCGCTCCACATCAACGAGCGCGGAGATCGCGTGGCCAGCACCA 468

QY 220 G 220

Db 469 G 469

#### RESULT 9

BQ691308

LOCUS

DEFINITION BQ691308 911 bp mRNA linear EST 15-JUL-2002

AGENCOURT 8345892 NIH\_MGC\_110 Homo sapiens cDNA clone IMAGE:6250859

5' mRNA sequence.

ACCESION BQ691308

VERSION BQ691308.1 GI:21816624

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

NIH-MGC http://mgi.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished

Contact: Robert Strausberg, Ph.D.

Email: sgapbs@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LUCM394 row: j column: 12

High quality sequence stop: 677.

Location/Qualifiers

1..911

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:6250859"

/tissue\_type="ductal carcinoma, cell line"

#### FEATURES

source

Query Match

Best Local Similarity

Matches

Conservative

Mismatches

Indels

Gaps

0;

0;

0;

0;

/lab\_host="DH10B (phage-resistant)"

/clone\_lib="NIH\_MGC\_110"

/note="Organ: pancreas; Vector: pOTB7; Site: 1: XhoI;

Site: 2: EcoRI; cDNA made by oligo-dT priming.

Directionally cloned into EcoRI/XhoI sites using the

following 5' adaptor: GGCACGAG(G). Library constructed by

Ling Hong in the laboratory of Gerald M. Rubin (University

of California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies).

Note: this is a NIH\_MGC Library."

216 a 234 c 255 g 201 t

5 others

Query Match 13.0%; Score 34.6; DB 13; Length 911;

Best Local Similarity 55.4%; Pred. No. 4.7;

Matches 67; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 100 TCAGGAACGTCGCCGCTTGCCTTGGCGGTCCACCAATGCCGTAAATAATCAGCAACCGCAG 159

Db 403 TAAGGACCGCGCGCTCTCACCGTGGCGTCACCAACCGTGGCGAGCTCCATCTCTCG 462

QY 160 CAATGAAGGATAGCATCACTTACCGCACTGTGGTAAAGCGGTGACCAATGCTAT 219

Db 463 CGAGACCGACTGCGCGCTCCACATCAACGAGCGCGGAGATCGCGTGGCCAGCACCA 522

QY 220 G 220

Db 523 G 523

#### RESULT 10

BH422048/c

LOCUS

DEFINITION BH422048 811 bp DNA linear GSS 12-DEC-2001

BOGLS36TF BOGL Brassica oleracea genomic clone BOGLS36, genomic

survey sequence.

ACCESION BH422048

VERSION BH422048.1 GI:17607776

KEYWORDS GSS.

SOURCE Brassica oleracea

ORGANISM Brassica oleracea

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

SpERMATOPHYTES; Magnoliophyta; eudicotyledons; core eudicots; rosids

; eurosids II; Brassicales; Brassicaceae; Brassica.

1 (bases 1 to 811)

Town, C.D., Van Aken, S., Utterback, T., Koo, H. and Fraser, C.M.

Whole genome shotgun sequencing of Brassica oleracea

Unpublished

Other GSSs: BOGLS36TF

Contact: Chris Town

TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA.

Tel: 301-838-3523

Fax: 301-838-0208

Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.

Seq primer: TF

Class: sheared ends.

Location/Qualifiers

1..811

/organism="Brassica oleracea"

/mol\_type="genomic DNA"

/strain="T01000DH3"

/db\_xref="taxon:3712"

/clone="BOGLS36"

/clone\_lib="BOGL"

/note="Vector: pHOS1; Site 1: BstXI; 2-3 kb sheared

genomic DNA inserted into pHOS1 using BstXI linkers"

308 a 179 c 143 g 181 t

Query Match

Best Local Similarity

Matches

Conservative

Mismatches

Indels

Gaps

0;

0;

0;

0;

0;

0;

```

QY 6 TGAACACGCATCATGTTTCGGCTCGTTCGTTGGTTCGATTAGCAGGTTCGGCT 65
Db 325 TGTGGAACATCATGTTTCCTGACGCCATCAATGCACCTTGCATTAGACCCCTGGCGGT 266
QY 66 CAATCAATAATGTAAACGGTTCGACACAGAACTTCAGGAACGTCGCGGTTCGCTTG 125
Db 265 CTAATTTCTAGTGTAGTGTAGTAGACAAATTTACGTTGTGTCATGAATTCCTCAA 206
QY 126 GCGTC 130
Db 205 GTGAC 201

RESULT 11
BH695730/c
LOCUS 345 bp DNA linear GSS 20-FEB-2002
DEFINITION BOMFEL4TR BO_2_3_KB Brassica oleracea genomic clone BOMFEL4,
genomic survey sequence.
ACCESSION BH695730
VERSION BH695730.1 GI:18768346
KEYWORDS GSS
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids II; Brassicales; Brassicaceae; Brassica.
AUTHORS 1 (bases 1 to 345)
Town, C.D., Van Aken, S., Utterback, T., Koo, H. and Fraser, C.M.
TITLE Whole genome shotgun sequencing of Brassica oleracea
JOURNAL Unpublished
COMMENT Other GSSs: BOMFEL4TF
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: ctown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TR
Class: sheared ends.

FEATURES
Source
Location/Qualifiers
1..345
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="TO100UDH3"
/db_xref="taxon:3712"
/clone="BOMFEL4"
/clone_lib="BO_2_3_KB"
/note="Vector: pHO51; Site 1: BstXI; 2-3 kb sheared
genomic DNA inserted into pHO51 using BstXI linkers"

BASE COUNT 73 a 78 c 93 g 101 t
ORIGIN
Query Match 12.5%; Score 33.4; DB 28; Length 345;
Best Local Similarity 52.5%; Pred. No. 7.2;
Matches 73; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 67 AATCAATATGTAACGTTTCGACCGAGAACTTCAGGAACGTCGCGGTTCGCTTGG 126
Db 339 AAACAAGAAGACTCCGTTTACACCAAAATCGCAGCGTGCATCATGTTGGGA 280
QY 127 CGTCACCAATCCGTAATAATCAGCAACCGCAGCAATGAAGGCATACGCATCACTTAC 186
Db 279 AGTAAATATCCCGATAGCAGAACTCGGACCCCAAGGCCAGCTACGAGCTTCC 220

QY 187 CGCAACTGTGGTAAGCGC 205
Db 219 GCCTAGCAATATTAGAGC 201

RESULT 12
BH165435
LOCUS 661 bp mRNA linear EST 09-JUL-2001
DEFINITION RE05337.5prime RE Drosophila melanogaster normalized Embryo pflc-1
Drosophila melanogaster cDNA clone RE05337 5, mRNA sequence.
ACCESSION BH165435
VERSION BH165435.1 GI:14631241
KEYWORDS EST.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
REFERENCE Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
AUTHORS 1 (bases 1 to 661)
Stapleton, M., Brokstein, P., Hong, L., Tyler, D., Berman, B., Carlson
, J., Champe, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E., George
, R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G., Misra, S.,
Mungall, C.J., Nunoo, J., Pacleb, J., Paragas, V., Park, S.,
Phouanavong, S., Wan, K., Yu, C., Lewis, S.E., Celnikier, S. and Rubin
, G.M.
TITLE BDGP/HMI RE Drosophila EST Project
JOURNAL Unpublished
COMMENT Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
hit genomic: AB003518; arm:3L [18805396,19084925]
estimated-cyto:75E4-76A1: 04/11/2001
Plate: RE.53 row: D column: 1
High quality sequence stop: 561.
Location/Qualifiers
1..661
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
/clone="RE05337"
/sex="male and female"
/dev_stage="0-24 hours mixed stage embryonic"
/lab_host="DHS-alpha Tona"
/clone_lib="RE Drosophila melanogaster normalized Embryo
pflc-1"
/note="Organ: embryo; Vector: pFlc1; Site 1: XhoI; Site 2:
BamHI; Library was kindly generated by Piero Carninci at
the RIKEN. The library was normalized and excised using
Cre recombinase. Plasmid cDNA library."

BASE COUNT 242 a 141 c 138 g 140 t
ORIGIN
Query Match 12.5%; Score 33.4; DB 12; Length 661;
Best Local Similarity 49.2%; Pred. No. 9.8;
Matches 88; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 77 GTACCGTTTCGACCAAACTTCAGAACTGCGCGTTTCGCTTGGCGTCACCAAT 136
Db 2 GTTACCGTTTCGCTCGCTAACGACAGCAAAATCGAGAGCTGCTTCGTCATTAA 61
QY 137 GCGCTAAATAATCAGCAACCGCAGCAATGAAGGCATACGCATCAACTTTACGCAACTGTG 196
Db 82 GCCAAACCAATCAAAACCAACAGACCGAATATATGAAGTAACAGTTTCAGAAACAGAC 121
QY 197 GGTAGCGCGTGCAACATGCTATGTTACCATGTATATCAGCACAATCGCGCTTACCCT 255
Db 122 GAAATTTCTCCGCAACTCGCGGTAAAGTGGGAATAAAACAAATTGGATATACAAT 180

RESULT 13
BH605016
LOCUS 730 bp DNA linear GSS 15-DEC-2001
DEFINITION BOHOA02TF BOHO Brassica oleracea genomic clone BOHOA02, genomic
survey sequence.
ACCESSION BH605016
VERSION BH605016.1 GI:17857462
KEYWORDS GSS.

```



```

Plate: LLCMI72 row: m column: 23
High quality sequence stop: 222.
Location/Qualifiers
1 .883
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon.9606"
/clone="IMAGE:350186"
/tissue_type="neuroblastoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 19"
/note="Organ: Brain; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."
Note: This is a NIH MGC Library."
BASE COUNT      188 a   277 C   281 G   137 T
ORIGIN
Query Match          12.5%; Score 33.4; DB 10; Length 883;
Best Local Similarity 62.7%; Pred. No. 11;
Matches 52; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 185 ACCGCCAATCTGGGTAAAGCGCGTGACCATTGTCATTACCACTGTAATCAGCACATCG 244
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
328 ACCCCAGGCGCCCTAAGGCGATCAGCCCCCATGATACCCGCCCATCAGCACATTCG 269

QY 245 GC GTTACCACTTC CGATGCAATT 267
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
268 GCCATTTCACAACCAATGCAACT 246

RESULT 15
LOCUS CNS02GGP/c
FEATURES             Location/Qualifiers
     DEFINITION       Tetraodon nigroviridis genome survey sequence T7 end of clone
                        137021 of library G from Tetraodon nigroviridis, genomic survey
                        sequence.
     ACCESSION         AL156306
     VERSION           AL196306.1 GI:7834456
     KEYWORDS           GSS; genome survey sequence.
     SOURCE            Tetraodon nigroviridis
     ORGANISM          Tetraodon nigroviridis
                        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                        Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
                        Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
                        Tetraodontoidea; Tetraodontidae; Tetraodon.
REFERENCE
  1 Roest Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
    Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
    Saurin,W. and Weissenbach,J
    Estimate of human gene number provided by genome-wide analysis
    using Tetraodon nigroviridis DNA sequence
    Nat. Genet. 25 (2), 235-238 (2000)
    20296633
    10835645
REFERENCE
  2 Roest Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
    Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
    Saurin,W., Bernot,A. and Weissenbach,J.
    Characterization and repeat analysis of the compact genome of the
    freshwater pufferfish Tetraodon nigroviridis
    Genome Res. 10 (7), 939-949 (2000)
    20359837
    10839143
REFERENCE
  3 (bases 1 to 962)
    Direct Submission
    Genoscope.
    Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
    BP 191 91006 EVRY Cedex - FRANCE [E-mail : seqref@genoscope.cns.fr]
JOURNAL
MEDLINE
PUBMED
TITLE
AUTHORS

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 15, 2003, 00:06:39 ; Search time 41.5874 Seconds  
(without alignments)  
2833.774 Million cell updates/sec

Title: US-09-928-457-89

Perfect score: 267

Sequence: 1 AATTATGACACGACATCA.....TTACACTTCGATGCAATT 267

Scoring table: IDENTITY NJC

Gapop 10.0 , Gapext 1.0

Searched: 563978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA:  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	34.6	13.0	1796	2	US-08-911-445-17
2	34.6	13.0	1796	3	US-09-182-983-17
3	34.6	13.0	1845	2	US-08-911-445-5
4	34.6	13.0	1845	3	US-09-182-983-5
5	34.6	13.0	2046	2	US-08-911-445-6
6	34.6	13.0	2046	3	US-09-182-983-6
7	30.4	11.4	1466	4	US-08-620-332D-914
8	29.4	11.0	1455	3	US-08-276-531-33
9	29	10.9	3947	3	US-08-975-762-47
10	29	10.9	3947	3	US-08-295-028-47
11	29	10.9	3947	4	US-09-106-582-47
12	28.2	10.6	1665	3	US-09-292-768-69
13	28.2	10.6	1764	3	US-08-268-532-17
14	28	10.5	8040	1	US-08-596-291-1
15	28	10.5	8040	3	US-09-100-804-1
16	28	10.5	8043	5	PCT-US94-09943-1
17	28	10.5	8119	3	US-09-290-640-45
18	28	10.5	4403765	3	US-09-103-840A-2
19	28	10.5	4411529	3	US-09-103-840A-1
20	27.8	10.4	2430	1	US-08-082-368-1
21	27.8	10.4	7986	4	US-08-647-452A-1
22	27.6	10.3	1103	1	US-08-181-271A-3
23	27.6	10.3	1103	1	US-08-449-315-3
24	27.6	10.3	1103	1	US-08-444-803-3
25	27.6	10.3	1103	1	US-08-449-043-3
26	27.6	10.3	1103	1	US-08-456-265A-3
27	27.6	10.3	1103	1	US-08-455-416-3

28	27.6	10.3	1103	1	US-08-455-244-3	Sequence 3, Appl
29	27.6	10.3	1103	1	US-08-454-876-3	Sequence 3, Appl
30	27.6	10.3	1103	2	US-08-457-264-3	Sequence 3, Appl
31	27.6	10.3	1103	2	US-08-456-262-3	Sequence 3, Appl
32	27.6	10.3	1103	2	US-08-456-240-3	Sequence 3, Appl
33	27.6	10.3	1103	2	US-08-455-736-3	Sequence 3, Appl
34	27.6	10.3	1103	2	US-08-971-217-3	Sequence 3, Appl
35	27.6	10.3	1103	3	US-08-329-799-38	Sequence 38, Appl
36	27.6	10.3	1103	3	US-09-350-600-3	Sequence 3, Appl
37	27.6	10.3	1224	4	US-09-615-192A-363	Sequence 363, Appl
38	27.6	10.3	2200	4	US-09-504-358-27	Sequence 27, Appl
39	27.6	10.3	2200	4	US-09-554-314-27	Sequence 27, Appl
40	27.6	10.3	11471	4	US-09-504-358-16	Sequence 16, Appl
41	27.6	10.3	11471	4	US-09-954-314-16	Sequence 16, Appl
42	27.6	10.3	12124	1	US-08-181-271A-36	Sequence 36, Appl
43	27.6	10.3	12124	1	US-08-449-315-36	Sequence 36, Appl
44	27.6	10.3	12124	1	US-08-444-803-36	Sequence 36, Appl
45	27.6	10.3	12124	1	US-08-449-043-36	Sequence 36, Appl

ALIGNMENTS

RESULT 1  
US-08-911-445-17  
; Sequence 17, Application US/08911445  
; Patent No. 5876713  
; GENERAL INFORMATION:  
; APPLICANT: TAKEDA CHEMICAL INDUSTRIES, LTD.  
; TITLE OF INVENTION: NOVEL GLUTAMINE:FRUCTOSE-6-PHOSPHATE  
; TITLE OF INVENTION: AMIDOTRANSFERASE, ITS PRODUCTION AND USE  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FETHERSTONHAUGH & Co.  
; STREET: P.O. BOX 2999, STATION D  
; CITY: OTTAWA  
; STATE: ONTARIO  
; COUNTRY: CANADA  
; ZIP: K1P 5Y6  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/911,445  
; FILING DATE: 12-AUG-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 213944-1996  
; FILING DATE: 13-AUG-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: FETHERSTONHAUGH & Co.,  
; REFERENCE/DOCKET NUMBER: 28605-30  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (613)-232-2486  
; TELEFAX: (613)-232-8440  
; INFORMATION FOR SEQ ID NO: 17:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1796 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; US-08-911-445-17

Query Match 13.0%; Score 34.6; DB 2; Length 1796;  
Best Local Similarity 55.4%; Pred. No. 0.043;  
Matches 67; Conservative 0; Mismatches 54; Indels 0; Gaps 0;  
QY 100 TCAGGAACGTGGCGCTTTCCTTGGCGTCCATGCCGTAAAAATCAGAACCCGAG 159  
DB 1319 TAAGGACCGGGCGCTCTCACCGTGGGCGTCCAAACCCGCGGAGCTCATCTCTCG 1378

QY 160 CAATGAAGGCATACGCATCACTTTACCGCAACTGTGGTAAGCGGTGACCAATGCTAT 219  
Db 1379 CGAGACCGACTGCGCGCTCTCACATCAACGAGCGCGGATCGCGTGGCCAGCACCA 1438  
QY 220 G 220  
Db 1439 G 1439

RESULT 2  
US-09-182-983-17  
; Sequence 17, Application US/09182983  
; Patent No. 6207431  
; GENERAL INFORMATION:  
; APPLICANT: TAKEDA CHEMICAL INDUSTRIES, LTD.  
; TITLE OF INVENTION: NOVEL GLUTAMINE:FRUCTOSE-6-PHOSPHATE  
; TITLE OF INVENTION: AMIDOTRANSFERASE, ITS PRODUCTION AND USE  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FETHERSTONHAUGH & Co.  
; STREET: P.O. BOX 2999, STATION D  
; CITY: OTTAWA  
; STATE: ONTARIO  
; COUNTRY: CANADA  
; ZIP: K1P 5Y6  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09182,983  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/911,445  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: FETHERSTONHAUGH & Co.,  
; REFERENCE/DOCKET NUMBER: 28605-30  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (613)-232-2486  
; TELEFAX: (613)-232-8440  
; INFORMATION FOR SEQ ID NO: 17:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1796 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
US-09-182-983-17

Query Match 13.0%; Score 34.6; DB 3; Length 1796;  
Best Local Similarity 55.4%; Pred. No. 0.043; Indels 0; Gaps 0;  
Matches 67; Conservative 0; Mismatches 54; Indels 0; Gaps 0;  
QY 100 TCAGGAAGCTGCGCGTGTGCTTGGCGGTCAACCAATGCCGTAAATAATCAGCAACCGCAG 159  
Db 1319 TAAGGACCGCGCGTCTCACCGTGGCGTCAACCAACACCGTGGCGAGTCCATCTCTCG 1378  
QY 160 CAATGAAGGCATACGCATCACTTTACCGCAACTGTGGTAAGCGGTGACCAATGCTAT 219  
Db 1379 CGAGACCGACTGCGCGCTCTCACATCAACGAGCGCGGATCGCGTGGCCAGCACCA 1438  
QY 220 G 220  
Db 1439 G 1439

RESULT 3  
US-08-911-445-5  
; Sequence 5, Application US/08911445

; Patent No. 5876713  
; GENERAL INFORMATION:  
; APPLICANT: TAKEDA CHEMICAL INDUSTRIES, LTD.  
; TITLE OF INVENTION: NOVEL GLUTAMINE:FRUCTOSE-6-PHOSPHATE  
; TITLE OF INVENTION: AMIDOTRANSFERASE, ITS PRODUCTION AND USE  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FETHERSTONHAUGH & Co.  
; STREET: P.O. BOX 2999, STATION D  
; CITY: OTTAWA  
; STATE: ONTARIO  
; COUNTRY: CANADA  
; ZIP: K1P 5Y6  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/911,445  
; FILING DATE: 12-AUG-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 213944-1996  
; FILING DATE: 13-AUG-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: FETHERSTONHAUGH & Co.,  
; REFERENCE/DOCKET NUMBER: 28605-30  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (613)-232-2486  
; TELEFAX: (613)-232-8440  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1845 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
US-08-911-445-5

Query Match 13.0%; Score 34.6; DB 2; Length 1845;  
Best Local Similarity 55.4%; Pred. No. 0.043; Indels 0; Gaps 0;  
Matches 67; Conservative 0; Mismatches 54; Indels 0; Gaps 0;  
QY 100 TCAGGAAGCTGCGCGTGTGCTTGGCGGTCAACCAATGCCGTAAATAATCAGCAACCGCAG 159  
Db 1311 TAAGGACCGCGCGTCTCACCGTGGCGTCAACCAACACCGTGGCGAGTCCATCTCTCG 1370  
QY 160 CAATGAAGGCATACGCATCACTTTACCGCAACTGTGGTAAGCGGTGACCAATGCTAT 219  
Db 1371 CGAGACCGACTGCGCGTCTCACATCAACGAGCGCGGAGATCGCGTGGCCAGCACCA 1430  
QY 220 G 220  
Db 1431 G 1431

RESULT 4  
US-09-182-983-5  
; Sequence 5, Application US/09182983  
; Patent No. 6207431  
; GENERAL INFORMATION:  
; APPLICANT: TAKEDA CHEMICAL INDUSTRIES, LTD.  
; TITLE OF INVENTION: NOVEL GLUTAMINE:FRUCTOSE-6-PHOSPHATE  
; TITLE OF INVENTION: AMIDOTRANSFERASE, ITS PRODUCTION AND USE  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FETHERSTONHAUGH & Co.  
; STREET: P.O. BOX 2999, STATION D  
; CITY: OTTAWA  
; STATE: ONTARIO  
; COUNTRY: CANADA  
; ZIP: K1P 5Y6

```

/ NAME: FETHERSTONHAUGH & Co.,
/ REFERENCE/DOCKET NUMBER: 28605-30
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (613)-232-2486
/ TELEFAX: (613)-232-8440
/ INFORMATION FOR SEQ ID NO: 6:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 2046 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: linear
/ MOLECULE TYPE: cDNA
/
US-08-911-443-6

Query Match 13.0%; Score 34.6; DB 2; Length 2046;
Best Local Similarity 55.4%; Pred.No.0.046;
Matches 67; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 100 TCAGGAACGTGCGCGCTTTGCCTTGGCGGTCAACCAATGTTTACCGCAACTGTGGGTAAACGGGTGACCAATGCTAT 150
Db 1311 TAAGGACCGCGCGCTCTCACCGTGGCGGTCAACCAACCGGTGGGCGAGATCGGCGTGGCCAGCACAA 1430
QY 160 CAATGAAGGCATACGCATCAACTTTACCGCAACTGTGGGTAAACGGGTGACCAATGCTAT 219
Db 1371 CGAGACCGACTGCGGCGTCCACATCAACGAGCGGCGAGATCGGCGTGGCCAGCACAA 1430
QY 220 G 220
Db 1431 G 1431

RESULT 6
US-09-182-983-6
/ Sequence 6, Application US/09182983
/ Patent No. 6207431
/ GENERAL INFORMATION:
/ APPLICANT: TAKEDA CHEMICAL INDUSTRIES, LTD.
/ TITLE OF INVENTION: NOVEL GLUTAMINE:FRUCTOSE-6-PHOSPHATE
/ TITLE OF INVENTION: AMIDOTRANSFERASE, ITS PRODUCTION AND USE
/ NUMBER OF SEQUENCES: 25
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: FETHERSTONHAUGH & Co.
/ STREET: P.O. BOX 2999, STATION D
/ CITY: OTTAWA
/ STATE: ONTARIO
/ COUNTRY: CANADA
/ ZIP: K1P 5Y6
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent In Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/182,983
/ FILING DATE:
/ CLASSIFICATION:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 08/911,445
/ FILING DATE:
/ ATTORNEY/AGENT INFORMATION:
/ NAME: FETHERSTONHAUGH & Co.,
/ REFERENCE/DOCKET NUMBER: 28605-30
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (613)-232-2486
/ TELEFAX: (613)-232-8440
/ INFORMATION FOR SEQ ID NO: 6:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 2046 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: linear
/ MOLECULE TYPE: cDNA
/
US-09-182-983-6

```

Query Match 13.0%; Score 34.6; DB 3; Length 2046;  
 Best Local Similarity 55.4%; Pred. No. 0.046;  
 Matches 67; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 100 TCAGGACGTCGCGGTTGGCGTACCAATGCCGTAAATACGCAACCCAG 159  
 DB 1311 TAAGACCGCGCGGCTTCACGTTGGCGGTACCAACACCGTGGCGAGCTCCATCTCTCG 1370

QY 160 CAATGAAGGCATAGCATCAACTTACCGCAACTGTGGTAAGCGGTGACCAATGCTAT 219  
 DB 1371 CGAGACCGACTGCGGCGTCCACATCAACGAGCGCGGAGATCGGCGTCCAGCACAA 1430

QY 220 G 220  
 DB 1431 G 1431

## RESULT 7

US-09-620-312D-914  
 ; Sequence 914, Application US/09620312D

; Patent No. 6569662

; GENERAL INFORMATION:

; APPLICANT: Tang, Y. Tom

; APPLICANT: Liu, Chenghua

; APPLICANT: Asundi, Vinod

; APPLICANT: Zhang, Jie

; APPLICANT: Ren, Feiyen

; APPLICANT: Chen, Rui-hong

; APPLICANT: Zhao, Qing A.

; APPLICANT: Wehrman, Tom

; APPLICANT: Xue, Aidong J.

; APPLICANT: Yang, Yongshong

; APPLICANT: Wang, Jian-Rui

; APPLICANT: Zhou, Ping

; APPLICANT: Ma, Yunqing

; APPLICANT: Wang, Dunrui

; APPLICANT: Wang, Zhiwei

; APPLICANT: John Tillinghast

; APPLICANT: Dmanac, Radole T.

; TITLE OF INVENTION: No. 6569662a1 Nucleic Acids and

; FILE REFERENCE: Polypeptides

; CURRENT APPLICATION NUMBER: US/09/620.312D

; CURRENT FILING DATE: 2000-07-19

; PRIOR APPLICATION NUMBER: 09/552,317

; PRIOR FILING DATE: 2000-04-25

; PRIOR APPLICATION NUMBER: 09/488,725

; PRIOR FILING DATE: 2000-01-21

; NUMBER OF SEQ ID NOS: 1105

; SOFTWARE: pt FL\_genes Version 1.0

; SEQ ID NO 914

; LENGTH: 1466

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (779)..(1213)

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (1)..(1466)

; OTHER INFORMATION: n = a,t,c or g

US-09-620-312D-914

Query Match 11.4%; Score 30.4; DB 4; Length 1466;  
 Best Local Similarity 57.3%; Pred. No. 1.1;  
 Matches 55; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 89 GACGAGAACTTCAGGACGTCGCGGTTTCCTTGGCGTCCACCAATGCGGTAAATC 148

DB 478 GAGGAGAGCTTCGGAACAGCGTCCGCCAAGGCGCTCAGTTTCCATTATGTC 537

QY 149 AGCAACCCGCAATGAAGGCATAGCATCAACTTT 184

DB 538 AACACGACCTTGAAGACATGAGGAGCAACACGTT 573

## RESULT 8

US-09-276-531-33

; Sequence 33, Application US/09276531

; Patent No. 6183968

; GENERAL INFORMATION:

; APPLICANT: Bandman, Olga

; APPLICANT: Lal, Preeti

; APPLICANT: Hillman, Jennifer L.

; APPLICANT: Yue, Henry

; APPLICANT: Reddy, Roopa

; APPLICANT: Guegler, Karl J.

; APPLICANT: Baughn, Mariah R.

; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF GENES ENCODING

; NUMBER OF SEQUENCES: 134

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.

; STREET: 3174 PORTER DRIVE

; CITY: PALO ALTO

; STATE: CALIFORNIA

; COUNTRY: USA

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/276.531

; FILING DATE: Herewith

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 60/079,677

; FILING DATE: March 27, 1998

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: Lynn E. Murry, Ph.D.

; REGISTRATION NUMBER: 42,918

; REFERENCE/DOCKET NUMBER: PA-0008 US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (650) 855-0555

; TELEFAX: (650) 845-4166

; INFORMATION FOR SEQ ID NO: 33:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1455 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; IMMEDIATE SOURCE:

; LIBRARY: BRAITUT01

; CLONE: 746308

US-09-276-531-33

Query Match 11.0%; Score 29.4; DB 3; Length 1455;  
 Best Local Similarity 56.8%; Pred. No. 2.4;  
 Matches 54; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 18 TCATCGTTTCGCGTCTCGTTGGTGGATTAGCAGTTGCGGTCATCAATAATG 77

DB 94 TCGAGTTCTGCAATCTGTGCGCGGTGAGCAATCGAGCGCTGGCTCACTCAATAAG 153

QY 78 TAACCGTTTCGACCAAGAACTTCAGGAACGTGCC 112

DB 154 GATCGGGGCCAACACGACATTCACAGTAGTTCC 188

## RESULT 9

US-08-975-762-47/c

; Sequence 47, Application US/08975762

```
Patent No. 6207169
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND
NUMBER OF SEQUENCES: 73
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/975,762
FILING DATE: 21-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.439
TELEPHONE: 206-622-4900
TELEFAX: 206-682-6031
INFORMATION FOR SEQ ID NO: 47:
SEQUENCE CHARACTERISTICS:
LENGTH: 3947 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-975-762-47

Query Match 10.9%; Score 29; DB 3; Length 3947;
Best Local Similarity 49.7%; Pred. No. 5.2;
Matches 74; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

Qy 71 AATAATGTAACCGTTTCGACAGCAAACTTCAGGAACGTGCGCGTTTGCCTTGGGGGTC 130
Db 2148 AATAGTCTGCTGATTTCTTATAAGTTAAATATAGCTCGTGGCGAATTCGGC 2089

Qy 131 ACCAATGCGGTAAATACGCAACCCGACGCAATGAAGGCATACGCATCAACTTTACCGCA 190
Db 2088 ACGAGTATTAGGGGCGAGAAATTTTCACAGCATTCGCAAACTGCAAAAGTCTTTACCGGA 2029

Qy 191 ACTGTGGGTAAGCGGTGACCAATGCTAT 219
Db 2028 AGTTTGGCAGAGCAGCAGCAAGTTAT 2000

RESULT 10
US-09-295-028-47/c
Sequence 47, Application US/09295028
Patent No. 6277381
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond L.
APPLICANT: McNeill, Patricia D.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
AND TREATMENT OF EHRlichia INFECTION
FILE REFERENCE: 210121.439C4
CURRENT APPLICATION NUMBER: US/09/295,028
CURRENT FILING DATE: 1999-04-20
NUMBER OF SEQ ID NOS: 85
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 47
```

```
LENGTH: 3947
TYPE: DNA
ORGANISM: Ehrlichia sp.
US-09-295-028-47

Query Match 10.9%; Score 29; DB 3; Length 3947;
Best Local Similarity 49.7%; Pred. No. 5.2;
Matches 74; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

Qy 71 AATAATGTAACCGTTTCGACAGCAAACTTCAGGAACGTGCGCGTTTGCCTTGGGGGTC 130
Db 2148 AATAGTCTGCTGATTTCTTATAAGTTAAATATAGCTCGTGGCGAATTCGGC 2089

Qy 131 ACCAATGCGGTAAATACGCAACCCGACGCAATGAAGGCATACGCATCAACTTTACCGCA 190
Db 2088 ACGAGTATTAGGGGCGAGAAATTTTCACAGCATTCGCAAACTGCAAAAGTCTTTACCGGA 2029

Qy 191 ACTGTGGGTAAGCGGTGACCAATGCTAT 219
Db 2028 AGTTTGGCAGAGCAGCAGCAAGTTAT 2000

RESULT 11
US-09-106-582-47/c
Sequence 47, Application US/09106582
Patent No. 6306402
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND
THERAPY
NUMBER OF SEQUENCES: 73
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/106,582
FILING DATE: 29-JUN-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.439C2
TELEPHONE: 206-622-4900
TELEFAX: 206-682-6031
INFORMATION FOR SEQ ID NO: 47:
SEQUENCE CHARACTERISTICS:
LENGTH: 3947 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-106-582-47

Query Match 10.9%; Score 29; DB 4; Length 3947;
Best Local Similarity 49.7%; Pred. No. 5.2;
Matches 74; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

Qy 71 AATAATGTAACCGTTTCGACAGCAAACTTCAGGAACGTGCGCGTTTGCCTTGGGGGTC 130
Db 2148 AATAGTCTGCTGATTTCTTATAAGTTAAATATAGCTCGTGGCGAATTCGGC 2089

Qy 131 ACCAATGCGGTAAATACGCAACCCGACGCAATGAAGGCATACGCATCAACTTTACCGCA 190
```

Db 2088 ACAGTATTAGGGGAGAGAAATTTTACAGCATTCGCAAACTGCAAAAGTCTTTTACCGGA 2029  
Qy 191 ACTGTGGGTAAAGCGGTGACCAATCTAT 219  
Db 2028 AGTTTGGCAAGAGCAGCAAGGTTAT 2000

## RESULT 12

US-09-292-768-69  
; Sequence 69, Application US/09292768  
; Patent No. 6194185  
; GENERAL INFORMATION:  
; APPLICANT: Croteau, Rodney B  
; APPLICANT: Lupien, Shari L  
; APPLICANT: Karp, Frank  
; TITLE OF INVENTION: RECOMBINANT MATERIALS AND METHODS FOR THE PRODUCTION OF  
; FILE OF INVENTION: LIMONENE HYDROXYLASES  
; FILE REFERENCE: wslr13463  
; CURRENT APPLICATION NUMBER: US/09/292,768  
; CURRENT FILING DATE: 1999-04-14  
; EARLIER APPLICATION NUMBER: 08/881,784  
; EARLIER FILING DATE: 1997-06-24  
; NUMBER OF SEQ ID NOS: 70  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 69  
; LENGTH: 1665  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:  
; OTHER INFORMATION: computer-generated nucleic acid sequence encoding  
; OTHER INFORMATION: limonene-3-hydroxylase  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: (1)..(1665)  
; OTHER INFORMATION: computer-generated nucleic acid sequence encoding  
; OTHER INFORMATION: limonene-3-hydroxylase protein variant  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (19)..(1518)  
US-09-292-768-69

Query Match 10.6%; Score 28.2; DB 3; Length 1665;  
Best Local Similarity 47.9%; Pred. No. 6.6;  
Matches 81; Conservative 0; Mismatches 88; Indels 0; Gaps 0;  
Qy 61 CGGCTCATCATATATGTAAACGGTTTCGACCAAGAACTTCAGGAAGTGCCTGGCGTTTC 120  
Db 753 CGAGCCCATCTCGAGGCCATCGTGGAGCAGCAAGCTCAAGAGAGCGCGGAGTTGG 812  
Qy 121 CTGGCGGTCCCAATGCGGTAAATAATCAGCAACCGCAGCAATGAAGGCATACGCATCAA 180  
Db 813 CGCGAGAGACATTATTGAGTACTCTTTAGATCGAGAGATAGCCAGATCAAGTCCC 872  
Qy 191 CTTTACCCCACTGTGGGTAGCGCGTGAACCAATGCTATGTACCATGT 229  
Db 873 CATCACCACCAAGCCATCAAAAGCGTTCATCTTCGACACGTTCTCAGCG 921

## RESULT 13

US-09-269-592-17/c  
; Sequence 17, Application US/09269592  
; Patent No. 6235500  
; GENERAL INFORMATION:  
; APPLICANT: Sligar, Stephen G.  
; APPLICANT: Sanders, Kevin  
; TITLE OF INVENTION: OXYGEN-BINDING HEME PROTEINS INCORPORATING  
; FILE OF INVENTION: CIRCULARLY-PERMUDED GLOBINS  
; FILE REFERENCE: 22010-140  
; CURRENT APPLICATION NUMBER: US/09/269,592  
; CURRENT FILING DATE: 1999-06-28  
; PRIOR APPLICATION NUMBER: PCT/US97/17294

; PRIOR FILING DATE: 1997-09-26  
; PRIOR APPLICATION NUMBER: 60/026,831  
; PRIOR FILING DATE: 1996-09-26  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 17  
; LENGTH: 1764  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (34)..(1752)  
US-09-269-592-17

Query Match 10.6%; Score 28.2; DB 3; Length 1764;  
Best Local Similarity 57.3%; Pred. No. 6.8;  
Matches 51; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

Qy 124 GGGCGTCACCAATGCGGTAAATAATCAGCAACCGCAGCAATGAAGCATACGCATCAACTT 183  
Db 130 GAGCGTTCGGCATGTGCTCAACATGAGCAACAGCGTTAGTCAACCGTCAGCAACTTTT 71  
Qy 184 TACCGCAACTGTGGTAAAGCGGTGACCA 212  
Db 70 TACCATGACCTTAACCTGAGCGGAACCA 42

## RESULT 14

US-08-596-291-1  
; Sequence 1, Application US/08596291  
; Patent No. 5821075  
; GENERAL INFORMATION:  
; APPLICANT: GONEZ, LEONEL JORGE  
; APPLICANT: SARAS, JAN  
; APPLICANT: CLAESSON-WELSH, LENA  
; APPLICANT: HELDIN, CARL-HENRIK  
; TITLE OF INVENTION: PRIMARY STRUCTURE AND FUNCTIONAL  
; TITLE OF INVENTION: EXPRESSION OF NUCLEOTIDE SEQUENCES FOR NOVEL PROTEIN  
; TITLE OF INVENTION: TYROSINE PHOSPHATASES  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.  
; STREET: 600 ATLANTIC AVENUE  
; CITY: BOSTON  
; STATE: MASSACHUSETTS  
; COUNTRY: USA  
; ZIP: 02210  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/596,291  
; FILING DATE: 03-AUG-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/115,573  
; FILING DATE: 01-SEP-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: GATES, EDWARD R.  
; REGISTRATION NUMBER: 31,616  
; REFERENCE/DOCKET NUMBER: LQ461/7000  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/720-3500  
; TELEFAX: 617/720-2441  
; TELEX: 92-1742 EZEKIEL  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 8040 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear



MOLECULE TYPE: CDNA to mRNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE: HOMO SAPIENS  
FEATURE: HOMO SAPIENS  
NAME/KEY: CDS  
LOCATION: 78..7475  
US-08-596-291-1

Query Match 10.5%; Score 28; DB 1; Length 8040;  
Best Local Similarity 56.5%; Pred. No. 16;  
Matches 52; Conservative 0; Mismatches 40; Indels 0; Gaps 0;  
QY 15 GCATCATCGTTTCGGTCCTGCGTTGGCATTAGCAGTTGCGGCTCAATCAATA 74  
DB 220 GCTTCATCATTTCTCCATGGTCTGCTGTTGCTGCCATCTGTAGTGTGTCATTACAG 279  
QY 75 ATGTAACCGTTTCGACACAGAACTTCAGAA 106  
DB 280 ATGAAATATTTCCAAATCAGATCTTCGAGCA 311

RESULT 15

US-09-100-804-1  
Sequence 1, Application US/09100804  
Patent No. 6066472  
GENERAL INFORMATION:  
APPLICANT: GONZALEZ, LEONEL JORGE  
APPLICANT: SARAS, JAN  
APPLICANT: CLAESSON-WELSH, LENA  
APPLICANT: HELDIN, CARL-HENRIK  
TITLE OF INVENTION: PRIMARY STRUCTURE AND FUNCTIONAL  
TITLE OF INVENTION: EXPRESSION OF NUCLEOTIDE SEQUENCES FOR NOVEL PROTEIN  
TITLE OF INVENTION: TYROSINE PHOSPHATASES  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.  
STREET: 600 ATLANTIC AVENUE  
CITY: BOSTON  
STATE: MASSACHUSETTS  
COUNTRY: USA  
ZIP: 02210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/100.804  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/596,291  
FILING DATE: 09-AUG-1996  
APPLICATION NUMBER: US 08/115,573  
FILING DATE: 01-SEP-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/09943  
FILING DATE: 01-SEP-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: GATES, EDWARD R.  
REGISTRATION NUMBER: 31,616  
REFERENCE/DOCKET NUMBER: L0461/7003  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-720-3500  
TELEFAX: 617-720-2441  
TELEX:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 8040 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single

TOPOLOGY: linear  
MOLECULE TYPE: CDNA to mRNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE: HOMO SAPIENS  
FEATURE: HOMO SAPIENS  
NAME/KEY: CDS  
LOCATION: 78..7475  
US-09-100-804-1

Query Match 10.5%; Score 28; DB 3; Length 8040;  
Best Local Similarity 56.5%; Pred. No. 16;  
Matches 52; Conservative 0; Mismatches 40; Indels 0; Gaps 0;  
QY 15 GCATCATCGTTTCGGTCCTGCGTTGGCATTAGCAGTTGCGGCTCAATCAATA 74  
DB 220 GCTTCATCATTTCTCCATGGTCTGCTGTTGCTGCCATCTGTAGTGTGTCATTACAG 279  
QY 75 ATGTAACCGTTTCGACACAGAACTTCAGAA 106  
DB 280 ATGAAATATTTCCAAATCAGATCTTCGAGCA 311

Search completed: November 15, 2003, 08:09:32  
Job time : 46.5874 secs



Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	267	100.0	267	10	US-09-928-457-89	Sequence 89, Appl
2	34.6	13.0	1851	13	US-10-044-090-506	Sequence 506, Appl
3	34.6	13.0	1936	9	US-09-771-838A-17	Sequence 17, Appl
4	34.6	13.0	1845	9	US-09-771-838A-5	Sequence 5, Appl
5	34.6	13.0	2046	9	US-09-771-838A-6	Sequence 6, Appl
6	34.6	13.0	3023	9	US-09-079-892-4	Sequence 4, Appl
7	34.6	13.0	3048	13	US-10-044-090-685	Sequence 685, Appl
8	31.4	11.8	369	10	US-09-867-701-3070	Sequence 3070, Ap
9	31.2	11.7	1601042	12	US-10-027-632-59064	Sequence 59064, A
10	31.2	11.7	1601042	13	US-10-027-632-59064	Sequence 59064, A
11	30.4	11.4	1466	4	US-10-037-270-934	Sequence 914, App
12	30	11.2	74868	14	US-10-175-523-67	Sequence 67, Appl
13	29.4	11.0	1473	11	US-09-796-753-47	Sequence 47, Appl
14	29.4	11.0	1481	14	US-10-198-846-11017	Sequence 11017, A
15	29.4	11.0	1741	13	US-10-098-841-67	Sequence 67, Appl
16	29.4	11.0	2211	10	US-09-738-621-1688	Sequence 1688, Ap

```
QY 61 CGGCTCAATCAATATGTAACGGTTTCGACAGAAATTCAGAAAGTGCCTGGTTTC 120
DB 61 CGGCTCAATCAATATGTAACGGTTTCGACAGAAATTCAGAAAGTGCCTGGTTTC 120
QY 121 CTTGGCGTCAACCAATGCGGTAAATAATCAGCAACGCGAGCAATGAAGCATACGCATCA 180
DB 121 CTTGGCGTCAACCAATGCGGTAAATAATCAGCAACGCGAGCAATGAAGCATACGCATCA 180
QY 181 CTTTACCGCAACTGTGGTAAGCGGTGACCAATGCTATGTACCAAGTGAATCAGCACA 240
DB 181 CTTTACCGCAACTGTGGTAAGCGGTGACCAATGCTATGTACCAAGTGAATCAGCACA 240
QY 241 ATCGCGGTACCACTTCGATGCAATT 267
DB 241 ATCGCGGTACCACTTCGATGCAATT 267

RESULT 2
US-10-044-090-506
; Sequence 506, Application US/10044090
; Publication No. US20020137081A1
; GENERAL INFORMATION:
; APPLICANT: Olga Bandman
; TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
; FILE REFERENCE: PA-0028 US
; CURRENT APPLICATION NUMBER: US/10/044,090
; CURRENT FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: PERL Program
; SEQ ID NO 506
; LENGTH: 1651
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No. US20020137081A1 403121.11
US-10-044-090-506

Query Match 13.0%; Score 34.6; DB 13; Length 1651;
Best Local Similarity 55.4%; Pred. No. 0.2;
Matches 67; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 100 TCAGAACGTCGCGGTTTGCCTTGGCGTCAACCAATGCGGTAAATAATCAGCAACGCGAG 159
DB 207 TAAGGACCGCGCGGTTCTCACCGTGGCGTCAACCAACGCGTGGCGAGTCCATCTCTCG 266
QY 160 CAATCAAGGCATACGCATCACTTACCGCAACTGTGGTAAGCGCGTGACCAATGCTAT 219
DB 267 CGAGACCGACTGCGCGTCCATCAACGAGCGCGGAGATCGCGTGGCCAGCACCAA 326
QY 220 G 220
DB 327 G 327

RESULT 3
US-09-771-838A-17
; Sequence 17, Application US/09771838A
; Patent No. US20020028198A1
; GENERAL INFORMATION:
; APPLICANT: TAKEDA CHEMICAL INDUSTRIES, LTD.
; TITLE OF INVENTION: NOVEL GLUTAMINE:FRUCTOSE-6-PHOSPHATE
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FETHERSTONHAUGH & Co.
; STREET: P.O. BOX 2999, STATION D
; CITY: OTTAWA
; STATE: ONTARIO
; COUNTRY: CANADA
; ZIP: K1P 5Y6
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/771,838A
; FILING DATE: 29-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/182,983
; FILING DATE: 1997-08-12
; ATTORNEY/AGENT INFORMATION:
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/771,838A
; FILING DATE: 29-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/182,983
; FILING DATE: 1997-08-12
; ATTORNEY/AGENT INFORMATION:
```

```
QY 100 TCAGAACGTCGCGGTTTGCCTTGGCGTCAACCAATGCGGTAAATAATCAGCAACGCGAG 159
DB 1319 TAAGGACCGCGCGGTTCTCACCGTGGCGTCAACCAACGCGTGGCGAGTCCATCTCTCG 1378
QY 160 CAATCAAGGCATACGCATCACTTACCGCAACTGTGGTAAGCGCGTGACCAATGCTAT 219
DB 1379 CGAGACCGACTGCGCGTCCATCAACGAGCGCGGAGATCGCGTGGCCAGCACCAA 1438
QY 220 G 220
DB 1439 G 1439

RESULT 4
US-09-771-838A-5
; Sequence 5, Application US/09771838A
; Patent No. US20020028198A1
; GENERAL INFORMATION:
; APPLICANT: TAKEDA CHEMICAL INDUSTRIES, LTD.
; TITLE OF INVENTION: NOVEL GLUTAMINE:FRUCTOSE-6-PHOSPHATE
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FETHERSTONHAUGH & Co.
; STREET: P.O. BOX 2999, STATION D
; CITY: OTTAWA
; STATE: ONTARIO
; COUNTRY: CANADA
; ZIP: K1P 5Y6
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/771,838A
; FILING DATE: 29-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/182,983
; FILING DATE: 1997-08-12
; ATTORNEY/AGENT INFORMATION:
```

NAME: FETHERSTONHAUGH & Co.,  
REFERENCE/DOCKET NUMBER: 28605-30  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (613)-232-2486  
TELEFAX: (613)-232-8440  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1845 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
SEQUENCE DESCRIPTION: SEQ ID NO: 5:  
US-09-771-838A-5

Query Match 13.0%; Score 34.6; DB 9; Length 1845;  
Best Local Similarity 55.4%; Pred. No. 0.22;  
Matches 67; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 100 TCAGGAAGTGGCGGTTTGCTTGGGGGTCCACCAATGCCGTAAATAATCAGCAACCGCAG 159  
Db 1311 TAAGGACCGGGGCTCTCACGGTGGGGGTCCACCAACCGTGGCGAGCTCCATCTCTCG 1370

QY 160 CAATGAAGCATACGATCACTTACCGCACTGTGGTAAAGCGCGTGACCAATGCTAT 219  
Db 1371 CGAGACCACTGGCGGCTCCACATCAACGAGGCGCGAGATCGGCGTGGCCAGCACAA 1430

QY 220 G 220  
Db 1431 G 1431

RESULT 5  
US-09-771-838A-6  
Sequence 6, Application US/09771838A  
Patent No. US20020028198A1  
GENERAL INFORMATION:  
APPLICANT: TAKEDA CHEMICAL INDUSTRIES, LTD.  
TITLE OF INVENTION: NOVEL GLUTAMINE-FRUCTOSE-6-PHOSPHATE  
AMIDOTRANSFERASE, ITS PRODUCTION AND USE  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FETHERSTONHAUGH & Co.  
STREET: P.O. BOX 2999, STATION D  
CITY: OTTAWA  
STATE: ONTARIO  
COUNTRY: CANADA  
ZIP: K1P 5Y6  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/771,838A  
FILING DATE: 29-Jan-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/182,983  
FILING DATE: 1997-08-12  
ATTORNEY/AGENT INFORMATION:  
NAME: FETHERSTONHAUGH & Co.,  
REFERENCE/DOCKET NUMBER: 28605-30  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (613)-232-2486  
TELEFAX: (613)-232-8440  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2046 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA

SEQUENCE DESCRIPTION: SEQ ID NO: 6:  
US-09-771-838A-6

Query Match 13.0%; Score 34.6; DB 9; Length 2046;  
Best Local Similarity 55.4%; Pred. No. 0.23;  
Matches 67; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 100 TCAGGAAGTGGCGGTTTGCTTGGGGGTCCACCAATGCCGTAAATAATCAGCAACCGCAG 159  
Db 1311 TAAGGACCGGGGCTCTCACGGTGGGGGTCCACCAACCGTGGCGAGCTCCATCTCTCG 1370

QY 160 CAATGAAGCATACGATCACTTACCGCACTGTGGTAAAGCGCGTGACCAATGCTAT 219  
Db 1371 CGAGACCACTGGCGGCTCCACATCAACGAGGCGCGAGATCGGCGTGGCCAGCACAA 1430

QY 220 G 220  
Db 1431 G 1431

RESULT 6  
US-09-079-892-4  
Sequence 4, Application US/09079892  
Patent No. US20020061301A1  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Lal, Freeta  
APPLICANT: Guegler, Karl J.  
APPLICANT: Gorgone, Gina  
APPLICANT: Corley, Neil C.  
APPLICANT: Patterson, Chandra  
APPLICANT: Baughn, Mariah R.  
TITLE OF INVENTION: HUMAN CARBOHYDRATE METABOLISM ENZYMES  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows  
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/079,892  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Cerrone, Michael C  
REGISTRATION NUMBER: 39,132  
REFERENCE/DOCKET NUMBER: PF-0524 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-855-0555  
TELEFAX: 650-855-0572  
TELEX:  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3023 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: SINTBST01  
CLONE: 1429011  
US-09-079-892-4

Query Match 13.0%; Score 34.6; DB 9; Length 3023;  
Best Local Similarity 55.4%; Pred. No. 0.28;  
Matches 67; Conservative 0; Mismatches 54; Indels 0; Gaps 0;  
QY 100 TCAGGAAGTCCCGGTTGGCGTCAACCAATGCCGTAATAATCAGCAACCGCAG 159  
DB 1409 TAAGACCGCGCGCTCTCACCGTGGCGTCAACCAACCGCTGGCGAGCTCCATCTCTCG 1468  
QY 160 CAATGAAGCATAGCATCAACTTTACCGCAACTGTGGTAAGCGCGTGACCAATGCTAT 219  
DB 1469 CGAGACCGACTGCGGCGTCCACATCAACGCGAGCGCGGAGATGGCGTGGCCAGCACC 1528  
QY 220 G 220  
DB 1529 G 1529

RESULT 7  
US-10-044-090-585  
; Sequence 685, Application US/10044090  
; Publication No. US20020137081A1  
; GENERAL INFORMATION:  
; APPLICANT: Olga Bandman  
; TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION  
; FILE REFERENCE: PA-0028 US  
; CURRENT APPLICATION NUMBER: US/10/044,090  
; CURRENT FILING DATE: 2002-01-09  
; NUMBER OF SEQ ID NOS: 850  
; SOFTWARE: PERL Program  
; SEQ ID NO 685  
; LENGTH: 3048  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; NAME/KEY: misc.feature  
; LOCATION: (1)...(10044090)  
; OTHER INFORMATION: Incyte ID No. US20020137081A1 403121.12  
US-10-044-090-585

Query Match 13.0%; Score 34.6; DB 13; Length 3048;  
Best Local Similarity 55.4%; Pred. No. 0.28; Indels 0; Gaps 0;  
Matches 67; Conservative 0; Mismatches 54; Indels 0; Gaps 0;  
QY 100 TCAGGAAGTCCCGGTTGGCGTCAACCAATGCCGTAATAATCAGCAACCGCAG 159  
DB 1429 TAAGACCGCGCGCTCTCACCGTGGCGTCAACCAACCGCTGGCGAGCTCCATCTCTCG 1488  
QY 160 CAATGAAGCATAGCATCAACTTTACCGCAACTGTGGTAAGCGCGTGACCAATGCTAT 219  
DB 1489 CGAGACCGACTGCGGCGTCCACATCAACGCGAGCGCGGAGATCGCGTGGCCAGCACC 1548  
QY 220 G 220  
DB 1549 G 1549

RESULT 8  
US-09-867-701-3070  
; Sequence 3070, Application US/09867701  
; Patent No. US20020132237A1  
; GENERAL INFORMATION:  
; APPLICANT: Aglate, Paul A.  
; APPLICANT: Jones, Robert  
; APPLICANT: Harlocker, Susan L.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER  
; FILE REFERENCE: 210121.497  
; CURRENT APPLICATION NUMBER: US/09/867,701  
; CURRENT FILING DATE: 2001-05-29  
; NUMBER OF SEQ ID NOS: 10912  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3070  
; LENGTH: 369  
; TYPE: DNA

; ORGANISM: Homo sapien  
US-09-867-701-3070  
Query Match 11.8%; Score 31.4; DB 10; Length 369;  
Best Local Similarity 54.9%; Pred. No. 1.3;  
Matches 62; Conservative 0; Mismatches 51; Indels 0; Gaps 0;  
QY 100 TCAGGAAGTCCCGGTTGGCGTCAACCAATGCCGTAATAATCAGCAACCGCAG 159  
DB 151 TAAGACCGCGCGCTCTCACCGTGGCGTCAACCAACCGCTGGCGAGCTCCATCTCTCG 210  
QY 160 CAATGAAGCATAGCATCAACTTTACCGCAACTGTGGTAAGCGCGTGACCA 212  
DB 211 CGAGACCGACTGCGGCGTCCACATCAACGCGAGCGCGGAGATCGCGTGGCCA 263

RESULT 9  
US-10-027-632-59064  
; Sequence 59064, Application US/10027632  
; Publication No. US20030204075A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; TITLE OF INVENTION: Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 59064  
; LENGTH: 1601042  
; TYPE: DNA  
; ORGANISM: Human  
; NAME/KEY: misc.feature  
; LOCATION: (1)...(1601042)  
; OTHER INFORMATION: n = A,T,C or G  
US-10-027-632-59064

Query Match 11.7%; Score 31.2; DB 12; Length 1601042;  
Best Local Similarity 64.3%; Pred. No. 94;  
Matches 45; Conservative 1; Mismatches 24; Indels 0; Gaps 0;  
QY 126 GGTCAACCAATCCGTAATAATCAGCAACCGCAGCAATGAAGGCAATACGCATCACTTTA 185  
DB 1005504 GCATTAATTAATCAATATCAATCAAGCAAGAGCAGCTATCAGCTGATTGTTAACAATGA 1005563  
QY 186 CCGCACTGT 195  
DB 1005564 CCGCACTGT 1005573

RESULT 10  
US-10-027-632-59064  
; Sequence 59064, Application US/10027632  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; TITLE OF INVENTION: Polymorphisms in the Human Genome

```
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027.632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 59064
; LENGTH: 1601042
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(1601042)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-59064

Query Match      11.7%; Score 31.2; DB 13; Length 1601042;
Best Local Similarity 64.3%; Pred. No. 94;
Matches 45; Conservative 1; Mismatches 24; Indels 0; Gaps 0;

QY 126 CGGTCAACAATCCGCTAAATAATCAGCAACCCAGCAATGAGGCATAGGCATCAACTTAA 185
DB 1005504 GCATAATTAATGCCAATAATATCAGCAAYAGCGTATCAGCTGATTTTAAACAATAGA 1005553

QY 186 CGGCAACTGT 195
DB 1005564 CGGCACTGT 1005573

RESULT 11
US-10-037-270-914
; Sequence 914, Application US/10037270
; Publication No. US20030104529A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yungqing
; APPLICANT: Wang, Dairui
; APPLICANT: Wang, Zhiwei
; APPLICANT: Tillinghast, John
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US20030104529A1el Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 784CIP28
; CURRENT APPLICATION NUMBER: US/10/037,270
; CURRENT FILING DATE: 2002-01-04
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
```

```
; NUMBER OF SEQ ID NOS: 1104
; SOFTWARE: PC_FL_genes Version 1.0
; SEQ ID NO 914
; LENGTH: 1466
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (779)..(1213)
; NAME/KEY: misc feature
; LOCATION: (1)..(1466)
; OTHER INFORMATION: n = a,t,c or g
US-10-037-270-914

Query Match      11.4%; Score 30.4; DB 14; Length 1466;
Best Local Similarity 57.3%; Pred. No. 5.8;
Matches 55; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 89 GACCAGAAACTTCAGGAAGCTGCGGGTTTGCTTTGGCGGTCCACCAATCCGTAATAATC 148
DB 478 GAGGAGAAGCTTCGGAACACAGGCTGCTCCCCCAAGGGCGCTCAGTTTCCATTATGTC 537

QY 149 AGCAACCGCAGCAATGAAGCATACCATCACTTT 184
DB 538 AACACCAGCCTTGAAGACATGAGGAGCAACAACGTT 573

RESULT 12
US-10-175-523-67
; Sequence 67, Application US/10175523
; Publication No. US20030096264A1
; GENERAL INFORMATION:
; APPLICANT: Brockman, Jeffrey
; APPLICANT: Evans, David
; APPLICANT: Hook, Derek
; APPLICANT: Klimczak, Leszek
; APPLICANT: Laeng, Pascal
; APPLICANT: Palfreyman, Michael
; APPLICANT: Rajan, Prithi
; TITLE OF INVENTION: MULTI-PARAMETER HIGH THROUGHPUT SCREENING ASSAYS (MPHTS)
; FILE REFERENCE: 3235/1J795-US3
; CURRENT APPLICATION NUMBER: US/10/175,523
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: US 60/299,151
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: US 60/317,828
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: US 60/325,150
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/333,047
; PRIOR FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: US 60/349,936
; PRIOR FILING DATE: 2002-01-18
; PRIOR APPLICATION NUMBER: US 60/361,834
; PRIOR FILING DATE: 2002-03-04
; NUMBER OF SEQ ID NOS: 197
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 67
; LENGTH: 74868
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-175-523-67

Query Match      11.2%; Score 30; DB 14; Length 74868;
Best Local Similarity 61.5%; Pred. No. 56;
Matches 48; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 39 TTGCGTTGGCATTAGCAGGTTGCGGTCAATCAATATGTAACCGTTTCGACCAAGAAC 98
DB 13595 TAGGTTGGAGTTGAAGATATAGCTTATTCAGGATCAAGCCATTGACGCGGCAA 13654

QY 99 TTCAGGAACGTGCCGCT 116
```

Db 13655 TTTAGAAAAGAGGCTGT 13672

RESULT 13

US-09-796-753-47

Sequence 47, Application US/09796753

Publication No. US20030027998A1

GENERAL INFORMATION:

APPLICANT: McCarthy, Sean A.

TITLE OF INVENTION: SECRETED PROTEINS AND USES THEREOF

FILE REFERENCE: 7853-227-999

CURRENT FILING DATE: 2001-03-01

CURRENT APPLICATION NUMBER: US/09/796,753

PRIOR APPLICATION NUMBER: 09/183,175

PRIOR FILING DATE: 1998-10-30

PRIOR APPLICATION NUMBER: 09/223,094

PRIOR FILING DATE: 1998-12-30

PRIOR APPLICATION NUMBER: 09/223,546

PRIOR FILING DATE: 1998-12-30

PRIOR APPLICATION NUMBER: 09/224,246

PRIOR FILING DATE: 1998-12-30

PRIOR APPLICATION NUMBER: 09/259,388

PRIOR FILING DATE: 1999-02-26

PRIOR APPLICATION NUMBER: 60/122,458

PRIOR FILING DATE: 1999-03-01

PRIOR APPLICATION NUMBER: 09/312,359

PRIOR FILING DATE: 1999-05-14

PRIOR APPLICATION NUMBER: 09/336,536

PRIOR FILING DATE: 1999-06-18

PRIOR APPLICATION NUMBER: 09/342,887

PRIOR FILING DATE: 1999-06-29

PRIOR APPLICATION NUMBER: 09/345,464

PRIOR FILING DATE: 1999-06-30

PRIOR APPLICATION NUMBER: 09/365,164

PRIOR FILING DATE: 1999-07-30

PRIOR APPLICATION NUMBER: 09/399,723

PRIOR FILING DATE: 1999-09-20

PRIOR APPLICATION NUMBER: 09/409,634

PRIOR FILING DATE: 1999-09-30

PRIOR APPLICATION NUMBER: 09/471,179

PRIOR FILING DATE: 1999-12-23

PRIOR APPLICATION NUMBER: 09/474,071

PRIOR FILING DATE: 1999-12-29

PRIOR APPLICATION NUMBER: 09/474,072

PRIOR FILING DATE: 1999-12-29

PRIOR APPLICATION NUMBER: 09/514,010

PRIOR FILING DATE: 2000-02-25

PRIOR APPLICATION NUMBER: 09/516,745

PRIOR FILING DATE: 2000-03-01

PRIOR APPLICATION NUMBER: 09/572,002

PRIOR FILING DATE: 2000-05-14

PRIOR APPLICATION NUMBER: 09/597,993

PRIOR FILING DATE: 2000-06-19

PRIOR APPLICATION NUMBER: 09/599,596

PRIOR FILING DATE: 2000-06-22

PRIOR APPLICATION NUMBER: 09/630,334

PRIOR FILING DATE: 2000-07-31

PRIOR APPLICATION NUMBER: 09/606,565

PRIOR FILING DATE: 2000-06-29

PRIOR APPLICATION NUMBER: 09/606,317

PRIOR FILING DATE: 2000-06-29

PRIOR APPLICATION NUMBER: 09/665,666

PRIOR FILING DATE: 2000-09-20

PRIOR APPLICATION NUMBER: 09/677,751

PRIOR FILING DATE: 2000-09-30

NUMBER OF SEQ ID NOS: 162

SEQ ID NO 47

LENGTH: 1473

TYPE: DNA

ORGANISM: Homo sapiens

US-09-796-753-47

Query Match 11.0%; Score 29.4; DB 11; Length 1473;

Best Local Similarity 56.8%; Pred. No. 13;

Matches 54; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 18 TCATCGTTTCGGCTCGGTTGGTTGGCATTAGCAGGTTGGCGCTCAATCAATAATG 77

Db 93 TCGCAGTTTCTGCATTCTGTGGGCGGTGAGCAATCGCAGCGCTGGCTCAGTCAATAAG 152

QY 78 TAACCGTTTCCGACAGAGAACTTCAGGAACGTGCC 112

Db 153 GATCCGGGCCCAACACGACATTCACAGTAGTTCC 187

RESULT 14

US-10-198-846-11017

Sequence 11017, Application US/10198846

Publication No. US20030099974A1

GENERAL INFORMATION:

APPLICANT: Lillie, James

APPLICANT: Xu, Yongyao

APPLICANT: Wang, Youzhen

APPLICANT: Steinmann, Kathleen

TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS

TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND

TITLE OF INVENTION: THERAPY OF BREAST CANCER

FILE REFERENCE: MRI-049

CURRENT APPLICATION NUMBER: US/10/198,846

CURRENT FILING DATE: 2002-07-18

PRIOR APPLICATION NUMBER: 60/306,220

PRIOR FILING DATE: 2001-07-18

NUMBER OF SEQ ID NOS: 14084

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 11017

LENGTH: 1481

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: misc feature

LOCATION: 1469, 1470, 1471, 1472, 1473, 1474, 1475, 1476, 1477, 1478,

LOCATION: 1479, 1480, 1481

OTHER INFORMATION: n = A,T,C or G

US-10-198-846-11017

Query Match 11.0%; Score 29.4; DB 14; Length 1481;

Best Local Similarity 56.8%; Pred. No. 13;

Matches 54; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 18 TCATCGTTTCGGCTCGGTTGGTTGGCATTAGCAGGTTGGCGCTCAATCAATAATG 77

Db 100 TCGCAGTTTCTGCATTCTGTGGGCGGTGAGCAATCGCAGCGCTGGCTCAGTCAATAAG 159

QY 78 TAACCGTTTCCGACAGAGAACTTCAGGAACGTGCC 112

Db 160 GATCCGGGCCCAACACGACATTCACAGTAGTTCC 194

RESULT 15

US-10-098-841-67

Sequence 67, Application US/10098841

Publication No. US20020197679A1

GENERAL INFORMATION:

APPLICANT: Tang, Y. Tom

APPLICANT: Liu, Chenghua

APPLICANT: Asundi, Vinod

APPLICANT: Xu, Chongjun

APPLICANT: Zhou, Ping

APPLICANT: Ma, Yunging

APPLICANT: Wang, Jian-Rui

APPLICANT: Zhao, Qing A.

APPLICANT: Ren, Feiyan

APPLICANT: Chen, Rui-hong

APPLICANT: Wang, Dunrui

APPLICANT: Wang, Zhiwei

APPLICANT: Wehrman, Tom



APPLICANT: Zhang, Jie  
APPLICANT: Qian, Xiaohong B.  
APPLICANT: Drmanac, Radoje T.  
TITLE OF INVENTION: No. US20020197679A1el Nucleic Acids and  
FILE REFERENCE: 784CIP2  
CURRENT APPLICATION NUMBER: US/10/098,841  
CURRENT FILING DATE: 2002-03-13  
PRIOR APPLICATION NUMBER: 09/598,042  
PRIOR FILING DATE: 2000-06-20  
PRIOR APPLICATION NUMBER: 09/552,317  
PRIOR FILING DATE: 2000-04-25  
PRIOR APPLICATION NUMBER: 09/488,725  
PRIOR FILING DATE: 2000-01-21  
NUMBER OF SEQ ID NOS: 331  
SOFTWARE: pt\_FL\_genes Version 1.0  
SEQ ID NO 67  
LENGTH: 1741  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (726)..(1412)  
NAME/KEY: misc feature  
LOCATION: (1)..(1741)  
OTHER INFORMATION: n = a,t,c or g  
US-10-098-841-67

Query Match 11.0%; Score 29.4; DB 13; Length 1741;  
Best Local Similarity 46.4%; Pred. No. 14;  
Matches 96; Conservative 0; Mismatches 111; Indels 0; Gaps 0;  
QY 57 GTTGGGGCTCAATCAATATGTAACGGTTTCGACACAGAACTTCAGGAACGTGCCGGT 116  
Db 1276 GTTGGGGCTCTACAGAGACCAGCCGCGCCGCGCCCTACCAGGCTCCCGGT 1335  
QY 117 TTGCTTTGGCGCTCACCATGCGTAAATAATCAGCAACCGCAATGAAGGCATACGCA 176  
Db 1336 CGGTATGCCCTCGCCACCTCGGACCAAGAGGGGAGAGAGCTTTGGCAATACGGCA 1395  
QY 177 TCACTTTACCGCAACTGTGGGTAAAGCGGTGACCAATGCTATGTTACAGTGAATCAG 236  
Db 1396 GAAAGCGCTACGTGTAGCAGCTCTGGCCGTGGGCGCGGTCTCTCCCACTGCCCAAG 1455  
QY 237 CACAATCGCGTTTACCACTTCGGATGC 263  
Db 1456 GAGAGGGGACCTGGCGGGGCCCAATC 1482

Search completed: November 15, 2003, 08:32:28  
Job time : 163.297 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: November 14, 2003, 23:56:24 ; Search time 1145.51 Seconds  
(without alignments)  
9241.609 Million cell updates/sec

Title: US-09-928-457-92  
Perfect score: 259  
Sequence: 1 AATTGTGTATATCAAGTAGG.....TACTTACTGCCAGCGAAATT 259

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_in.\*
- 18: em\_mu.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_mus.\*
- 34: em\_htg\_pln.\*
- 35: em\_htg\_rod.\*
- 36: em\_htg\_mam.\*
- 37: em\_htg\_vrt.\*
- 38: em\_sy.\*
- 39: em\_htgo\_hum.\*
- 40: em\_htgo\_mus.\*
- 41: em\_htgo\_other.\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	259	100.0	259	1	AF169468 Neisseria
2	259	100.0	259	6	A68921 Sequence 92
3	259	100.0	259	6	BD03021 DNA and s
C 4	255.8	98.8	349081	1	NMA222491
5	56.8	21.9	1275	6	A96274 Neisseria
6	52.6	20.3	1572	6	AX024062 Sequence 30
7	52.6	20.3	1575	6	A96276 Sequence 30
8	52.6	20.3	1575	6	AX024113 Sequence
9	50.4	19.5	15053	1	AE002550 Neisseria
10	50.4	19.5	172325	6	AX044035 Sequence
11	50.4	19.5	349980	6	AX044034 Sequence
C 12	41.2	15.9	13326	6	AX346615 Sequence
C 13	41	15.8	90284	8	T2K10
14	40.6	15.7	236159	2	AC125553
15	40.6	15.7	253253	2	AC126527 Rattus no
16	40	15.4	222039	2	AC107506 Rattus no
C 17	39.2	15.1	173138	5	AC108132 Homo sapi
C 18	39.2	15.1	180964	2	AC025846 Homo sapi
19	39.2	15.1	217384	2	AC023573 Homo sapi
C 20	39	15.1	327738	2	AC104922 Mus muscu
21	38.6	14.9	221866	2	AC114426 Mus muscu
22	38.4	14.8	91940	9	AL445674 Human DNA
23	38.2	14.7	15282	6	AX251836 Sequence
C 24	38.2	14.7	15282	6	AX348621 Sequence
C 25	38	14.7	6944	6	AX347467 Sequence
C 26	38	14.7	6944	6	AX349188 Sequence
C 27	38	14.7	6944	6	AX657867 Sequence
C 28	38	14.7	6944	6	AX659141 Sequence
C 29	38	14.7	16509	6	AX346222 Sequence
C 30	37.8	14.6	5552	6	AX346161 Sequence
31	37.6	14.5	1576	6	A96272 Sequence 30
32	37.6	14.5	1576	6	A96270 Sequence 30
33	37.2	14.4	94056	9	AL139192 Human DNA
34	37.2	14.4	146753	9	AC105753 Homo sapi
C 35	37.2	14.4	153615	2	AC111043 Mus muscu
36	37.2	14.4	181443	9	AC099775 Homo sapi
37	37	14.3	339	6	BD113032 EST and e
C 38	37	14.3	95976	9	AC068062 Homo sapi
C 39	37	14.3	349980	6	AX344560 Sequence
40	36.8	14.2	35301	8	AP002544 Arabidops
41	36.8	14.2	219340	10	AL844481 Mouse DNA
C 42	36.6	14.1	8576	6	AX347130 Sequence
C 43	36.4	14.1	2200	6	AX347092 Sequence
C 44	36.4	14.1	6239	6	AX251987 Sequence
C 45	36.4	14.1	6239	6	AX344383 Sequence

ALIGNMENTS

RESULT 1  
AF169468 AF169468 259 bp DNA linear BCT 09-AUG-2000  
LOCUS Neisseria meningitidis strain Z2491 clone Em045 unknown sequence.  
DEFINITION  
ACCESSION AF169468  
VERSION AF169468.1 GI:9754676  
KEYWORDS  
SOURCE Neisseria meningitidis  
ORGANISM Neisseria meningitidis  
Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
Neisseriaceae; Neisseria.  
REFERENCE 1 (bases 1 to 259)  
Perrin A., Nassif X. and Tinsley C.R.  
AUTHORS Identification of regions of the chromosome of Neisseria  
TITLE meningitidis and Neisseria gonorrhoeae which are specific to

Pred. No. is the number of results predicted by chance to have a

```

pathogenic Neisseriae
Unpublished
REFERENCE 2 (bases 1 to 259)
AUTHORS Perrin,A., Nassif,X. and Tinsley,C.R.
TITLE Direct Submission
JOURNAL Submitted (16-JUL-1999) Necker Medicine Faculty, INSERM U411, 156
rue de Vaugirard, Paris 75015, France
FEATURES
    source
        1..259
        /organism="Neisseria meningitidis"
        /mol_type="genomic DNA"
        /strain="Z2491"
        /db_xref="taxon:487"
        /clone="Em045"
BASE COUNT 81 a 71 c 34 g 73 t
ORIGIN
Query Match 100.0%; Score 259; DB 1; Length 259;
Best Local Similarity 100.0%; Pred. No. 4.7e-65;
Matches 259; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AATTGTGTATATCAAGTAGGATGGCGCATTTATGCTGACCTACAAACCAAAACCAACCT 60
DB 1 AATTGTGTATATCAAGTAGGATGGCGCATTTATGCTGACCTACAAACCAAAACCAACCT 60
QY 61 ACCACCCCTTAATCAACTCCACAAACCCCTCTTTCAGACAAACCTCGTTTTTTGAAAAACAATC 120
DB 61 ACCACCCCTTAATCAACTCCACAAACCCCTCTTTCAGACAAACCTCGTTTTTTGAAAAACAATC 120
QY 121 TGTAAACAGATAACTGCTGAAGATACCGTTGCCGAGCCCAAAACCCGTAAGTCAACTT 180
DB 121 TGTAAACAGATAACTGCTGAAGATACCGTTGCCGAGCCCAAAACCCGTAAGTCAACTT 180
QY 181 TTATTGTGAACCTTCCCATTTATGAGAAATCCCTTTTCGTCTCTTCTTGTATTGTCCTCCT 240
DB 181 TTATTGTGAACCTTCCCATTTATGAGAAATCCCTTTTCGTCTCTTCTTGTATTGTCCTCCT 240
QY 241 ACTTACTGCCAGCGAAATT 259
DB 241 ACTTACTGCCAGCGAAATT 259

RESULT 3
LOCUS BD063021
DEFINITION DNA and specific proteins or peptides of the Neisseria meningitidis
species bacteria, method for obtaining them and their biological
applications.
ACCESSION BD063021
VERSION BD063021.1 GI:22608624
KEYWORDS JP 2001504684-A/83.
SOURCE unidentified
ORGANISM unidentified
REFERENCE 1 (bases 1 to 259)
AUTHORS Nassif,X., Tinsley,C., Achtman,M., Ruelle,J.L., Vinals,C. and
Merker,P.
TITLE DNA and specific proteins or peptides of the Neisseria meningitidis
species bacteria, method for obtaining them and their biological
applications
JOURNAL Patent: JP 2001504684-A 83 10-APR-2001;
INSTITUT NATIONAL DE LA SANTE ET DE LA RECHERCHE MEDICALE, MAX
PLANCK GESELLSCHAFT ZUR FORDERUNG DER WISSENSCHAFTEN EV BERLIN,
SMITHKLINE BEECHAM
COMMENT PN JP 2001504684-A/83
PD 10-APR-2001
PF 11-JUL-1997 JP 19980505685
PR 12-JUL-1996 FR 96/08768
PI XAVIER NASSIF, COLIN TINSLEY, MARK ACHTMAN, JEAN LOUIS RUELLE, PI
CARLA VINALS,
PI PETRA MERKER
PC Cl2N15/31,C07K14/22,C07K16/12,A61K39/095,C12Q1/68,G01N33/53 CC
Strandedness: Single;
CC Topology: Linear;
PH Key Location/Qualifiers.
FEATURES
    source
        1..259
        /organism="unidentified"
        /mol_type="genomic DNA"
        /db_xref="taxon:32644"
BASE COUNT 81 a 71 c 34 g 73 t
ORIGIN
Query Match 100.0%; Score 259; DB 6; Length 259;
Best Local Similarity 100.0%; Pred. No. 4.7e-65;
Matches 259; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AATTGTGTATATCAAGTAGGATGGCGCATTTATGCTGACCTACAAACCAAAACCAACCT 60
DB 1 AATTGTGTATATCAAGTAGGATGGCGCATTTATGCTGACCTACAAACCAAAACCAACCT 60
QY 61 ACCACCCCTTAATCAACTCCACAAACCCCTCTTTCAGACAAACCTCGTTTTTTGAAAAACAATC 120
DB 61 ACCACCCCTTAATCAACTCCACAAACCCCTCTTTCAGACAAACCTCGTTTTTTGAAAAACAATC 120
QY 121 TGTAAACAGATAACTGCTGAAGATACCGTTGCCGAGCCCAAAACCCGTAAGTCAACTT 180
DB 121 TGTAAACAGATAACTGCTGAAGATACCGTTGCCGAGCCCAAAACCCGTAAGTCAACTT 180
QY 181 TTATTGTGAACCTTCCCATTTATGAGAAATCCCTTTTCGTCTCTTCTTGTATTGTCCTCCT 240
DB 181 TTATTGTGAACCTTCCCATTTATGAGAAATCCCTTTTCGTCTCTTCTTGTATTGTCCTCCT 240
QY 241 ACTTACTGCCAGCGAAATT 259
DB 241 ACTTACTGCCAGCGAAATT 259

RESULT 2
LOCUS A68921
DEFINITION Sequence 92 from Patent WO9802547.
ACCESSION A68921
VERSION A68921.1 GI:4759840
KEYWORDS unidentified
SOURCE unidentified
ORGANISM unidentified
REFERENCE 1 (bases 1 to 259)
AUTHORS Nassif,X., Tinsley,C., Achtman,M., Ruelle,J., Vinals,C. and
Merker,P.
TITLE DNA AND SPECIFIC PROTEINS OR PEPTIDES OF THE NEISSERIA MENINGITIDIS
SPECIES BACTERIA, METHOD FOR OBTAINING THEM AND THEIR BIOLOGICAL
APPLICATIONS
JOURNAL Patent: WO 9802547-A 92 22-JAN-1998;
INST NAT SANTE RECH MED (FR)
COMMENT Other publication FR 2751000 19980116.
FEATURES
    source
        1..259
        /organism="unidentified"
        /mol_type="genomic DNA"
        /db_xref="taxon:32644"
BASE COUNT 81 a 71 c 34 g 73 t
ORIGIN
Query Match 100.0%; Score 259; DB 6; Length 259;
Best Local Similarity 100.0%; Pred. No. 4.7e-65;
Matches 259; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 121 TGTAAACAGATAACTGCTGAGATACCGTTGCGAGCCCAAAACCCGTAAGCACTT 180
Db 121 TGTAAACAGATAACTGCTGAGATACCGTTGCGAGCCCAAAACCCGTAAGCACTT 180
QY 181 TTATTGTGAACCTCCCAATTATGAGAAAATCCCTTTTCGTCCTCTTTCTGTAATTCGTCCTT 240
Db 181 TTATTGTGAACCTCCCAATTATGAGAAAATCCCTTTTCGTCCTCTTTCTGTAATTCGTCCTT 240
QY 241 ACTTACTGCCAGCGAAATT 259
Db 241 ACTTACTGCCAGCGAAATT 259

RESULT 4
NMA222491/C
LOCUS NMA222491 349061 bp DNA linear BCT 02-SEP-2002
DEFINITION Neisseria meningitidis serogroup A strain Z2491 complete genome;
segment 2/7.
ACCESSION AL162753 AL157959
VERSION AL162753.2 GI:7379120
SOURCE Neisseria meningitidis Z2491
ORGANISM Neisseria meningitidis Z2491
Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
Neisseriaceae; Neisseria.
REFERENCE 1 (bases 1 to 349061)
AUTHORS Parkhill,J., Achtman,M., James,K.D., Bentley,S.D., Churcher,C.,
Klee,S.R., Morelli,G., Basham,D., Brown,D., Chillingworth,T.,
Davies,R.M., Davis,P., Devlin,K., Feltwell,T., Hamlin,N.,
Holroyd,S., Jagers,K., Leather,S., Moule,S., Mungall,K.,
Quail,M.A., Rajandream,M.A., Rutherford,K.M., Simmonds,M.,
Skelton,J., Whitehead,S., Spratt,B.G. and Barrell,B.G.
Complete DNA sequence of a serogroup A strain of Neisseria
meningitidis Z2491
JOURNAL Nature 404 (6777), 502-506 (2000)
MEDLINE 20222556
PUBMED 10761919
REFERENCE 2 (bases 1 to 349061)
AUTHORS Parkhill,J.
TITLE Direct Submission
JOURNAL Submitted (30-MAR-2000) Submitted on behalf of the Neisseria
sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk
NOTES Details of N. meningitidis sequencing at the Sanger Centre are
available on the world wide Web.
(URL, http://www.sanger.ac.uk/Projects/N_meningitidis/).
FEATURES
source
1..349061
/organism="Neisseria meningitidis Z2491"
/mol_type="genomic DNA"
/strain="Z2491"
/db_xref="taxon:122587"
/notes="serogroup: A"
complement(24..206)
/notes="ATR repeat; hmms hit to HMM ATR (1 - 183), score:
310.39"
/label=ATR
209..212
220..681
/genes="NMA0368"
220..681
/genes="NMA0368"
/notes="NMA0368", probable integral membrane protein, len:
153 aa; contains four probable transmembrane domains"
/codon_start=1
/transl_table=11
/product="putative integral membrane protein"
/protein_id="CAB83669.1"
/db_xref="GI:7379121"
/db_xref="SPTREMBL:Q9JWH7"
/translation="MQQNRKSPFIVMLLVSVLWIAISLVNFAVLGNHGSMEGLTV

```

gene

CDS

```

LILGSI PASLDIRVCAVYVWLAAILVLLALRKVVPVHFAFWCLALVALVSKAVYV
DEAGNTSDIVRYGAGFYLYWAFVAFVAGTGTAGKXKKAASADGTXNDV"
671..1492
/genes="hemk"
671..1492
/genes="hemk"
/notes="NMA0369, hemk, Hemk protein, len: 273 aa; similar
to e.g. HEMK ECOLI P37186 HEMK protein (277 aa), fasta
scores; E(): 0.423% identity in 279 aa overlap. Contains
PS00092 N-6 Adenine-specific DNA methylases signature"
/codon_start=1
/transl_table=11
/product="Hemk protein"
/protein_id="CAB83670.1"
/db_xref="GI:7379122"
/db_xref="SPTREMBL:Q9JWH6"
/translation="MFDKWLGLSKLPKXNEARMMLQYVSEYTRVOLLTRGGEEMPDEV
KQRLDQRLKNGEPVAYILGAREFFGRFTVNPVSILPRPETHELVAVLARLPEN
GRWDLGTGSAVAVTALERPDFAVRASDI SPPALETARKAADLGRVAFYGSWF
DRLMPSGKWDIIVSNPPYINGDKHLSQGLRFPQIALTDFDGLSCLRTLAGAP
DRLAEGFLILLEHGFDQGAARGVLAENGPSGVETLPDLAGLDRLVTLCKYKHLK"
831..840
/genes="hemk"
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
1162..1171
/genes="hemk"
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
1187..1207
/genes="hemk"
/notes="PS00092 N-6 Adenine-specific DNA methylases
signature"
complement(1279..1288)
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
1565..1568
1578..1596
/genes="NMA0370"
/genes="NMA0370"
/notes="NMA0370, probable integral membrane protein, len:
462 aa; similar to hypothetical proteins e.g. Y125_HAEIN
P44640 hypothetical protein H10325 (450 aa), fasta scores;
E(): 0.498% identity in 464 aa overlap"
/codon_start=1
/transl_table=11
/product="putative integral membrane protein"
/protein_id="CAB83671.1"
/db_xref="GI:7379123"
/db_xref="SPTREMBL:Q9JWH5"
/translation="MNAVVAIVMVLVLSLRVHVLSLTIGAFVGGAVAGNPLQNI
DAAGVQAGIIPVFNKLEGAKIALSYAMLGAFAMATHSGLPQOLAGAVVRKLN
GMPDSVRSGEVGVKWLIIIVMGMSQNIPIPIAFIPVLPPLLVPNRLKIDR
RLIACVITFGVTVTFPLPGVGAIFLNLGNHSAAPOLDVKNVNMAMATPAL
CMLAGLLAFVHYKPRLYQSNADTADNADANRPOPSVRSIAAVALVAFPAIQL
MYESELVGLMGLFVFMGLVNNKDVFGEGIKMNAVGFIMIAQGPAAVMA
TGHIOPLVSSNAIFGNSKGMALVVLVGLVTMGISFSSTLIIAIIYVPLCVGL
GFSPLATVAIVGTAGALDAGSPASDSTLGTMLGNADQGDHIDRHSVPTFIHYNIP
LIIAGWIAAMVL"
2967..3191
/genes="slyx"
2967..3191
/genes="slyx"
/notes="NMA0371, slyX, slyX protein homolog, len: 74 aa;
similar to slyX HAEIN P44759 slyX protein homolog (73 aa),
fasta scores; E(): 0.0017, 33.8% identity in 74 aa
overlap, and slyX ECOLI P30857 slyX protein (72 aa), fasta
scores; E(): 0.24, 32.4% identity in 68 aa overlap"
/codon_start=1
/transl_table=11
/product="slyX protein homolog"
/protein_id="CAB83672.1"

```

misc\_feature

misc\_feature

misc\_feature

misc\_feature

RBS

gene

CDS

gene

CDS

```
/db_xref="GI:7379124"
/db_xref="SPRMBL:O9JWH4"
/translation="MDAVQSFPHRIEILETSALQEDVIAGLNAMVAELRQTLDLQQA
OURLYQMDRNPDAQEPYSLRDEIPPH"
complement(3271..3423)
/gene="NMA0372"
CDS
complement(3271..3423)
/gene="NMA0372"
note="NMA0372, unknown, questionable CDS, len: 50 aa"
/codon_start=1
/transl_table=1
/product="very hypothetical protein NMA0372"
/protein_id="CAB83673.1"
/db_xref="GI:7379125"
/db_xref="SPRMBL:O9JWH3"
/translation="MGRSLPDLGSGFFFLAVYEPVGFICLQKGTAYFRSGGKRFPHRI
KGIUSD"
3562..3571
misc_feature
/label="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(3586..4356)
/gene="thiF"
CDS
complement(3586..4356)
/gene="thiF"
/note="NMA0373, thiF, probable ThiF protein, len: 256 aa;
similar to e.g. THIF_ECOLI P30138 ThiF protein (251 aa),
fasta scores; E(): 0.43.1% identity in 246 aa overlap,
and MOBE_ECOLI P12282 molybdopterin biosynthesis MOBE
protein_(249 aa), fasta scores; E(): 0.43.9% identity in
244 aa overlap (note that N.m. does not have orthologs of
any other molybdopterin biosynthesis proteins). Contains
Pfam match to entry PF00899 ThiF_family, ThiF family"
/codon_start=1
/transl_table=1
/product="ThiF protein"
/protein_id="CAB83674.1"
/db_xref="GI:7379126"
/db_xref="SPRMBL:O9JWH2"
/translation="MTTEHDNDLFLRYSHILDEIGECQKLSAHLVVGCG
LGAAALFYLAAGIGTLTADSDTVLHNLQRFVDEGDKLTELADLRHIN
HTVDVRIKGLDGLTGVQADIVLDCDNDVATRVACVQAKTPLYSGAAVR
FEGGLAVRDLDPSPCYACLPDGGASDGLCSLFGVSPVLGIIGSTQAAEALKILL
DAGEPSGRILAVRALEGQYFDLPRNPECPVCGAER"
3787..3796
misc_feature
/label="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(3847..4257)
/gene="thiF"
/note="Pfam match to entry PF00899 ThiF_family, ThiF
family, score 186.60, E-value 4.1e-52"
complement(4418..4427)
/misc_feature
/label="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
4471..7224
/gene="ppc"
4471..7224
/gene="ppc"
/EC_number="4.1.1.31"
/note="NMA0374, ppc, phosphoenolpyruvate carboxylase, len:
917 aa; similar to many e.g. CAPP_RHOA O32483
phosphoenolpyruvate carboxylase (EC 4.1.1.31) (936 aa),
fasta scores; E(): 0.43.3% identity in 928 aa overlap.
Contains 2x pfam match to entry PF00311 PEPCase,
Phosphoenolpyruvate carboxylase, PS00017 ATP/GTP-binding
site motif A (P-loop), and PS00393 Phosphoenolpyruvate
carboxylase active site 2"
/codon_start=1
/transl_table=1
/product="phosphoenolpyruvate carboxylase"
/protein_id="CAB83675.1"
Query Match 98.8%; Score 255.8; DB 1; Length 349061;
Best Local Similarity 99.2%; Pred. No. 3.9e-64;
```

```
Matches 257; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 AATTGTGTATATCAAGTAGGATGGCATTATGCTGACCTACAAAACCAAAACAACT 60
Db 65981 AATTGTGTATATCAAGTAGGATGGCATTATGCTGACCTACAAAACCAAAACAACT 65922
QY 61 ACCACCCCTTAATCAACTCCACAAAACCCCTCTTCAGACAACCTCGTTTTTGAATAACAATC 120
Db 65921 ACCACCCCTTAATCAACTCCACAAAACCCCTCTTCAGACAACCTCGTTTTTGAATAACAATC 65862
QY 121 TGTAAACAGATAACTGTGTAAGATAACCGTTGCCGAGCCCAAAACCCGTAACCACTT 180
Db 65861 TGTAAACAGATAACTGTGTAAGATAACCGTTGCCGAGCCCAAAACCCGTAACCACTT 65802
QY 181 TTATTGTGAACCTTCCATTATGAGAAAATCCCTTTTGTGTCCTCTTCTGTATTCTGTCCT 240
Db 65801 TTATTGTGAACCTTCCATTATGAGAAAATCCCTTTTGTGTCCTCTTCTGTATTCTGTCCT 65742
QY 241 ACTTACTCCAGCGAAATT 259
Db 65741 ACTTACTCCAGCGAAATT 65723
RESULT 5
A96274 1275 bp DNA linear PAT 07-SEP-2000
LOCUS Sequence 307 from Patent WO9924578.
DEFINITION A96274
ACCESSION A96274
VERSION A96274.1 GI:5780029
KEYWORDS unidentified
SOURCE unidentified
ORGANISM unidentified
REFERENCE 1
AUTHORS Pizza, M., Scarlato, V., Rappuoli, R., Grandi, G. and Massignani, V.
TITLE Neisserial antigens
JOURNAL Patent: WO 9924578-A 307 20-MAY-1999;
PIZZA MARIAGRAZIA (IT); SCARLATO VINCENZO (IT); RAPPUOLI RINO (IT);
CHIRON SPA (IT); GRANDI GUIDO (IT); Massignani VEGA (IT)
FEATURES
Location/Qualifiers
1..1275
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"
BASE COUNT 300 a 304 c 319 g 352 t
ORIGIN
Query Match 21.9%; Score 56.8; DB 6; Length 1275;
Best Local Similarity 96.7%; Pred. No. 8.5e-06;
Matches 58; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 200 ATGAGAAATCCCTTTTGTGTCCTCTTCTGTATTCTGTCCTCTTCTGTCCTCTTCTGTCCTGCGAAATT 259
Db 1 ATGAGAAATCCCTTTTGTGTCCTCTTCTGTATTCTGTCCTCTTCTGTCCTCTTCTGTCCTGCGAAATT 60
RESULT 6
AX024062 1572 bp DNA linear PAT 15-SEP-2000
LOCUS Sequence 5 from Patent FR2785293.
DEFINITION AX024062
ACCESSION AX024062
VERSION AX024062.1 GI:10184374
KEYWORDS Neisseria meningitidis
SOURCE Neisseria meningitidis
ORGANISM Neisseria meningitidis
REFERENCE 1
AUTHORS Nassif, X., Tinsley, C., Autame, L., Perrin, A., Rokbi, B.,
Bouchardon, A. and Renaud, M.G.
JOURNAL Patent: FR 2785293-A 5 05-MAY-2000;
PASTEUR MERIEUX SERUMS VACC (FR)
FEATURES
Location/Qualifiers
```

```

source
1..1572
/organism="Neisseria meningitidis"
/mol_type="genomic DNA"
/db_xref="taxon:487"

CDS
1..1572
/notes="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAC08907.1"
/db_xref="GI:10184375"
/translation="MKKSLFVLFLYSSLLTASEIAYRVFGIETLPAKMAETFAITF
MIAALYLPARYKASRLIIAVFAFSNIANNVHYVQSMGTGINYKMLKEITREVGA
GASMLDKLWLPALWGVAEVLNFCSLAKFRKTHFSADILFALMLMIFVRSPTKQEH
GISEPKTYSRIKANYFSFGYFVGRVLPYQLFDLSKIPVFKQAPSKIGGSGNIIVLI
MGESAAHLKLFYGRGRTSPFLRLSQADFKPIVKQSYSGAFMTAVLSPFNVLPH
ANGLEQISGGDTNMFRLAKEQGYETFYFSAQAEENMAILNLIGKWDHLIQTOLGY
GNGDNMPDEKLLPLFPKINLQGRHFIVLHQGHAPYVALLQPDQVFBELIYDKDY
DNIHTKDDMIOTVFELQKQPDGNWLFAVTSDHGOYVRQDIYNQGTVPQSYLVLPLV
YSSNKAVQAAQNAQAPCEIAFHQQLSTFLHTLGYDMPVSGCREGSGVTGNLTGDAG
SLNTRDGAIEYVPO"
BASE COUNT      380 a   383 c   395 g   414 t
ORIGIN
Query Match      20.3%; Score 52.6; DB 6; Length 1572;
Best Local Similarity 93.2%; Pred. No. 0.00014;
Matches 55; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 200 ATGAGAAATCCCTTTTCGTCCTCTTCTGTTATTCGTCCTACTTACTGCGAGGAAAT 258
|||||
Db 1 ATGAAAAATCCCTTTTCGTCCTCTTCTGTTATTCATCCTACTTACCGCAGGAAAT 59
|||||

RESULT 7
A96276 A96276 1575 bp DNA linear PAT 07-SEP-2000
LOCUS Sequence 309 from Patent WO924578.
DEFINITION Neisserial antigens
ACCESSION A96276
VERSION A96276.1 GI:6780030
KEYWORDS unidentifed
SOURCE unidentifed
ORGANISM unclassified.
REFERENCE
1. Pizza, M., Scarlato, V., Rappuoli, R., Grandi, G. and Maignani, V.
Neisserial antigens
TITLE Patent: WO 924578-A 309 20-MAY-1999.
JOURNAL PIZZA MARIAGRAZIA (IT); SCARLATO VINCENZO (IT); RAPPUOLI RINO (IT);
CHIRON SPA (IT); GRANDI GUIDO (IT); MASNIGNANI VEGA (IT)
FEATURES
Location/Qualifiers
1..1575
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"
BASE COUNT      385 a   387 c   392 g   411 t
ORIGIN
Query Match      20.3%; Score 52.6; DB 6; Length 1575;
Best Local Similarity 93.2%; Pred. No. 0.00014;
Matches 55; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 200 ATGAGAAATCCCTTTTCGTCCTCTTCTGTTATTCGTCCTACTTACTGCGAGGAAAT 258
|||||
Db 1 ATGAAAAATCCCTTTTCGTCCTCTTCTGTTATTCATCCTACTTACCGCAGGAAAT 59
|||||

RESULT 8
AX024113 AX024113 1575 bp DNA linear PAT 15-SEP-2000
LOCUS Sequence 56 from Patent FR2785293.
DEFINITION Neisserial antigens
ACCESSION AX024113
VERSION AX024113.1 GI:10184424
KEYWORDS unidentifed
SOURCE Neisseria gonorrhoeae

```

---

```

ORGANISM
Neisseria gonorrhoeae
Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
Neisseriaceae; Neisseria.
REFERENCE
1. Nassif, X., Tinsley, C., Aujame, L., Perrin, A., Rokbi, B.,
Bouchardon, A. and Renaud, M.G.
JOURNAL Patent: FR 2785293-A 56 05-MAY-2000;
PASTEUR MERIEUX SERUMS VACC (FR)
FEATURES
Location/Qualifiers
1..1575
/organism="Neisseria gonorrhoeae"
/mol_type="genomic DNA"
/db_xref="taxon:485"
1..1575
/notes="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAC08932.1"
/db_xref="GI:10184425"
/translation="MKKSLFVLFLYSSLLTASEIAYRVFGIETLPAKMAETFAITF
MIAALYLPARYKASRLIIAVFAFSNIANNVHYVQSMGTGINYKMLKEITREVGA
GASMLDKLWLPALWGVAEVLNFCSLAKFRKTHFSADILFALMLMIFVRSPTKQEH
GISEPKTYSRIKANYFSFGYFVGRVLPYQLFDLSKIPVFKQAPSKIGGSGNIIVLI
MGESAAHLKLFYGRGRTSPFLRLSQADFKPIVKQSYSGAFMTAVLSPFNVLPH
ANGLEQISGGDTNMFRLAKEQGYETFYFSAQAEENMAILNLIGKWDHLIQTOLGY
GNGDNMPDEKLLPLFPKINLQGRHFIVLHQGHAPYVALLQPDQVFBELIYDKDY
DNIHTKDDMIOTVFELQKQPDGNWLFAVTSDHGOYVRQDIYNQGTVPQSYLVLPLV
LYSPDKAVQAAQNAQAPCEIAFHQQLSTFLHTLGYDMPVSGCREGSGVTGNLTGDA
GSLNTRDGAIEYVPO"
BASE COUNT      385 a   387 c   392 g   411 t
ORIGIN
Query Match      20.3%; Score 52.6; DB 6; Length 1575;
Best Local Similarity 93.2%; Pred. No. 0.00014;
Matches 55; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 200 ATGAGAAATCCCTTTTCGTCCTCTTCTGTTATTCGTCCTACTTACTGCGAGGAAAT 258
|||||
Db 1 ATGAAAAATCCCTTTTCGTCCTCTTCTGTTATTCATCCTACTTACCGCAGGAAAT 59
|||||

RESULT 9
AE002550 AE002550 15053 bp DNA linear BCT 25-MAY-2000
LOCUS Neisseria meningitidis serogroup B strain MC58 section 192 of 206
DEFINITION of the complete genome.
ACCESSION AE002550 AE002098
VERSION AE002550.2 GI:7413476
KEYWORDS Neisseria meningitidis MC58
SOURCE Neisseria meningitidis MC58
ORGANISM Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
Neisseriaceae; Neisseria.
1 (bases 1 to 15053)
Tettelin, H., Saunders, N.J., Heidelberg, J., Jeffries, A.C.,
Nelson, K.E., Eisen, J.A., Ketchum, K.A., Hood, D.W., Peden, J.F.,
Dodson, R.J., Nelson, W.C., Gwinn, M.L., DeBoy, R., Peterson, J.D.,
Hickey, E.K., Haft, D.H., Salzberg, S.L., White, O., Fleischmann, R.D.,
Dougherty, B.A., Mason, T., Ciecko, A., Parksey, D.S., Blair, E.,
Citron, H., Clark, E.B., Cotton, M.D., Utterback, T.R., Khouri, H.,
Qin, H., Vamathevan, J., Gill, J., Scarlato, V., Maignani, V.,
Pizza, M., Grandi, G., Sun, L., Smith, H.O., Fraser, C.M., Moxon, E.R.,
Rappuoli, R. and Venter, J.C.
Complete genome sequence of Neisseria meningitidis serogroup B
strain MC58
Science 287 (5459), 1809-1815 (2000)
JOURNAL MEDLINE
20175755
10710307
REFERENCE
2 (bases 1 to 15053)
Tettelin, H., Saunders, N.J., Heidelberg, J., Jeffries, A.C.,
Nelson, K.E., Eisen, J.A., Ketchum, K.A., Hood, D.W., Peden, J.F.,
Dodson, R.J., Nelson, W.C., Gwinn, M.L., DeBoy, R., Peterson, J.D.,
Hickey, E.K., Haft, D.H., Salzberg, S.L., White, O., Fleischmann, R.D.,

```







```

source
1. 13326
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"

BASE COUNT      3846 a 115 c 2866 g 6899 t
ORIGIN

Query Match      15.9%; Score 41.2; DB 6; Length 13326;
Best Local Similarity 54.7%; Pred. No. 0.31;
Matches 82; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 36 TGACCTACAAACCAAAACCACTACACCTTAATCACTCACAACACCTCTTCAGA 95
DB 8480 TTATCTATAAAACAAATACATATACCTCCCATCTCCCAACACACGACAACTCA 8421
QY 96 CAACCTCGTTTTTTGAAAAACAATCTGTAAACAGATAACTGCTGAAGAAATACCGTTGCCG 155
DB 8420 ACATACTCATCTTCTACAAAAAACTCCAAAATTTCTTTTCCACCACTCCCAATTAC 8361
QY 156 AGCCCCAAAACCGTACTGCAACTTTTATT 185
DB 8360 TACCACATAACACATACCTTAAATAATAAT 8331

RESULT 13
T2K10/c
LOCUS
DEFINITION
Arabidopsis thaliana chromosome 1 BAC T2K10 sequence, complete
sequence.
ACCESSION      AC005966
VERSION        AC005966.2 GI:12039260
KEYWORDS       HTG.
SOURCE         Arabidopsis thaliana (thale cress)
ORGANISM       Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 90284)
Vysotskaya,V.S., Schwartz,J.R., Yu,G., Toriumi,M., Lenz,C., Liu,S.,
Li,J., Kremenetskaia,I., Luros,J., Gonzalez,A., Altati,H.,
Araujo,R., Buchler,B., Chao,Q., Conn,L., Conway,A.B., Dunn,P.,
Hansen,M., Huizar,L., Kim,C., Palm,C., Rowley,D., Shinn,P.,
Walker,M., Davis,R.W., Ecker,J.R., Federspiel,N.A. and Theologis,A.
Arabidopsis thaliana chromosome 1 BAC T2K10 sequence
Unpublished
2 (bases 1 to 90284)
Theologis,A.
Direct Submission
Submitted (14-NOV-1998) Plant Gene Expression Center, 800 Buchanan
Street, Albany, CA 94710, USA
3 (bases 1 to 90284)
Theologis,A.
Direct Submission
Submitted (29-JAN-1999) Plant Gene Expression Center, 800 Buchanan
Street, Albany, CA 94710, USA
4 (bases 1 to 90284)
Theologis,A.
Direct Submission
Submitted (09-FEB-1999) Plant Gene Expression Center, 800 Buchanan
Street, Albany, CA 94710, USA
5 (bases 1 to 90284)
Theologis,A.
Direct Submission
Submitted (10-FEB-1999) Plant Gene Expression Center, 800 Buchanan
Street, Albany, CA 94710, USA
6 (bases 1 to 90284)
Theologis,A.
Direct Submission
Submitted (05-JAN-2001) Plant Gene Expression Center, 800 Buchanan
Street, Albany, CA 94710, USA
On Jan 5, 2001 this sequence version replaced gi:4204173.
This sequence is of BAC T2K10 from Arabidopsis thaliana chromosome

FEATURES
source
1. 90284
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"
/chromosome="1"
/clone="T2K10"
/complement(<1..871)
/gene="T2K10.1"
/complement(join(<1..199,543..871))
/gene="T2K10.1"
/note="Similar to gb|AF039182 probable aldo-keto reductase
from Fragaria x ananassa. This gene may be cut off. EST
gb|U74151 comes from this gene."
/codon_start=1
/evidence=not experimental
/protein_id="AADI4487.1"
/db_xref="GI:4249390"
/translation="MSLTITVPTLAIRSGPSGHSMPVLGCTAASPEPEPTMLKETVI
PAIKLGRHFTSPRYQTEETIGALAEAVSLGVRSEFFVTTKLWCADAHGLV
PAIKRSLKNLKLVDLYIIHWPVSSKPKYKFFIDEDDFNPMDFEVVWSEMSCQL
GLAKCIGVSNFSCCKL"
1405..2487
/gene="T2K10.2"
1405..2487
/gene="T2K10.2"
/note="Strong similarity to gi|2829864 F316.6 zinc
metalloproteinase homolog from Arabidopsis thaliana BAC
gb|AC002396. EST gb|226412 comes from this gene."
/codon_start=1
/evidence=not experimental
/protein_id="AADI4473.1"
/db_xref="GI:4249376"
/translation="WRTLLILIFFFTVNPISAKFTVNSSIPLQFLANATONAWET
FSKLACGHIGNINGLSKLQVFRFGYITTTGNCDDDFDVLQSAINTYQKNFLKV
TGLDSSLTROIIVPRCCNPDLQVSEMGKKILRTTEKYSFPFGKPRWPKRDLT
YAFAPQNLITEVKVEFSRATRWAVTPELNTSESLIRADIVIGFSGHGGGEFP
DAMGTLAHSPTFCMLHLDGDDWLSNGEISERILPVTWVLDLSSVAVHEIGLL
GLGHSVEDAIMFPAISGGRKRLAKDDIEGIQHLIGNNGNDGGGSKPSRESQSTG
GDSVRWRGWMVSISSATCFLISV"
8873..11295
/gene="T2K10.3"
1018(8873..9018,9166..9289,9521..9603,9732..9843,
9947..10051,10146..10196,10580..10688,10783..10985,
11074..11295)
/gene="T2K10.3"
/note="Similar to gi|2829865 F316.4 from Arabidopsis
thaliana BAC gb|AC002396."
/codon_start=1
/evidence=not experimental
/protein_id="AADI4474.1"
/db_xref="GI:4249377"
/translation="MATHSSRSSENKADGEDELRRNPYEVVLGIPSNSTDGEIKSAYR
RNALRHPDKNPDPVAAEMFKEVTFAYELVSDPENRLYDTTSEAVGPENEDLELD
LSSLGAVNTIFAALNKLGVQIKITVSANLIGALNGTITPLMVGQVSRKVKQS
AHFYGVLTTEEAQDLICKVHSSAKNFKLLYDQVNGGLSLAQALSRDPEITGF
FKRLDAFOPFITELKAGSHVFAVYGNDFKSVSYLTLEIFSSAPFNGEKSLRSTEAQ
IVSKETELKFEAEVHEVFAQETMAASKTGEVOELDELLKRNIEICAAVTFPPTKQ
GSKRSWSKKSSLLMFPREGEVAVRESGVKKKKWYNTQLRQDKKK"
complement(18173..>20461)
/gene="T2K10.4"
complement(join(18173..18222,18296..18347,18479..18631,
18720..19026,19139..19599,19676..19789,19878..20165,
20369..>20461))
/gene="T2K10.4"
/note="Identical to gb|AJ010471 mRNA for DEAD box RNA
helicase (RH22) from Arabidopsis thaliana. EST gb|Y11191
comes from this gene."
/codon_start=1

```

1. In order to facilitate the joining of overlapping clones in the future for creation of larger contigs, we provide overlap between overlapping submitted clones. The 3' end of this sequence overlaps by 2447 bp to the 5' end of the sequence of BAC T13D8.

/gene="T2K10.8"  
complement(join(41797..42015,42114..42274,42362..42531,  
42626..42684,42760..42846,42942..43038,43137..43321,  
43404..43463,43542..43617,43705..43820,43880..44056,  
44142..44239,44326..44452))  
/gene="T2K10.8"  
Note=Strong similarity to gi|3337350 Fl3P17.3 putative  
proteinase from Arabidopsis thaliana BAC gb|AC004481."  
Query Match 15.8%; Score 41; DB 8; Length 90284;  
Best Local Similarity 56.2%; Pred. No. 0.35;  
Matches 77; Conservative 0; Mismatches 60; Indels 0; Gaps 0;  
Qy 41 TACAAACCAAAACCACTACACCCCTTATCAACTCCACAAACCGTCTTACAGACAAC 100  
Db 48143 TAAATAAAAATATTTCAGTACAGCCAAATAATAAAAATTGCAAAACCACTTAAAGAGACT 4808  
Qy 101 TCGTTTTTTGAAAAACAATCTGTFAAACAGATACTGCTGAAGAATAACCGTTGCCGAGCCC 160  
Db 48083 GCGTCGATTATATAAATATTGTTGCAATTATTGAGGAAAATAACAGTTACGCGTTTC 4802  
Qy 161 CAAACCCGTAATGCAA 177  
Db 48023 ATAAAACTAAATAAAA 48007  
RESULT 14  
AC125553  
LOCUS  
DEFINITION  
Rattus norvegicus clone CH230-4H12, \*\*\* SEQUENCING IN PROGRESS \*\*\*,  
2 unordered pieces.  
AC125553  
AC125553.4 GI:30467820  
HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.  
KEYWORDS  
Rattus norvegicus (Norway rat)  
SOURCE  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
REFERENCE  
1 (bases 1 to 236159)  
Muzny,D,Marie,, Metzker,M, Lee,, Abramson,S., Adams,C., Alder,J.,  
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,  
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,  
Baldwin,D., Bandaraika,D., Barber,M., Barnstead,M., Benahmed,F.,  
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,  
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,  
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,  
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,  
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,  
Devila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,  
Delgado,O., Denison,S., Deranc,C., Ding,Y., Dinh,H., Divya,K.,  
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,  
Egan,A., Escoto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,  
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,  
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,  
Georgieff,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,  
Gunnaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K.,  
Harvey,Y., Hawlik,P., Hawes,A., Henderson,N., Hernandez,J.,  
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,  
Hollins,B., Howells,S., Huylk,S., Hume,J., Idlebird,D., Jackson,A.,  
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,  
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,  
Kwais,C., Kratt,C.L., Lebow,H., Levian,J., Lewis,L., Li,Z., Liu,J.,  
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,  
Lorenshew,L., Loulseghe,H., Lozado,R.J., Lu,X., Ma,J.,  
Maheshwari,M., Mahindartne,M., Mahmoud,M., Mallory,K., Mangum,A.,  
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,  
Mawney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,  
Milobavjevic,A., Miner,G., Mirja,E., Montemayor,J., Moore,S.,  
Morgan,M., Morris,K., Morris,S., Munitadasa,M., Murphy,M., Nair,L.,  
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,  
Nwaekolameh,O., Okwuonu,G., Olarnpungagoon,A., Pal,S., Parks,K.,  
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,

Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rokey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajda, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steinle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmami, K., Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, D., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Direct Submission  
Unpublished  
2 (bases 1 to 236159)  
Worley, K.C.

Direct Submission  
Submitted (29-JUN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 236159)

Direct Submission  
Rat Genome Sequencing Consortium.  
Submitted (09-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On May 9, 2003 this sequence version replaced gi:24941144.  
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequences are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
----- Project Information  
Center project name: GCHD  
Center clone name: CH230-4H12  
----- Summary Statistics  
Assembly program: Atlas  
Consensus quality: 227359 bases at least Q40  
Consensus quality: 228996 bases at least Q30  
Consensus quality: 229931 bases at least Q20  
Estimated insert size: 238976: sum-of-contigs estimation  
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

\* NOTE: Estimated insert size may differ from sequence length (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently consists of 2 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 234738: contig of 234738 bp in length  
234839 234838: gap of unknown length  
234839 236159: contig of 1321 bp in length.

Location/Qualifiers

1..236159  
/organism="Rattus norvegicus"  
/mol\_type="Genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-4H12"  
39881..42222  
/note="wgs\_contig"

BASE COUNT 65808 a 48158 c 47610 g 69047 t 5536 others  
ORIGIN

Query Match 15.7%; Score 40.6; DB 2; Length 236159;  
Best Local Similarity 52.0%; Pred. No. 0.46; Indels 0; Gaps 0;  
Matches 91; Conservative 0; Mismatches 84;

QY 50 AAAAAACAACCTTACCACTTAACTCAACCTTCCAGAACCTTCTTCAGACAACTCTTTT 109  
DB 171795 ACATATATAGATACAGAGTCAATATATAGAACCAAAATATATATCAATATCTTT 171854

QY 110 GAAAAACAATCTGTAAACAGATACCTGTGAAGATACCTTGCAGCCCAAAACCG 169  
DB 171855 TAAAGAGGAATTCACCACTGGTGTCTAGTATATATATCTTCCAGTCCCAACCATCCC 171914

QY 170 TACTGCAACTTTTATGTGAACCTTCCCAATATGAGAAATCCCTTTCTCTCTCT 224  
DB 171915 CTCTCAACTCTCTGTGTCTCTCCCTCTCAACCAAGCCCTTTTCATCTAT 171969

RESULT 15  
AC126527 253253 bp DNA linear HTG 13-MAY-2003  
LOCUS  
DEFINITION  
Rattus norvegicus clone CH230-157M5, WORKING DRAFT SEQUENCE, 3 unordered pieces.  
AC126527  
AC126527.5 GI:30573006  
HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
KEYWORDS  
Rattus norvegicus (Norway rat)  
SOURCE  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
1 (bases 1 to 253253)  
REFERENCE  
AUTHORS  
Muzny, D., Marie, Metzker, M., Lee, Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hughes, M., Hollins, B., Howell, S., Hulyk, S., Hume, J., Idebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshuwa, L., Loulseghe, H., Lozano, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangun, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokemele, O., Okwuonu, G., Olarpunsgoon, A., Pal, S., Parks, K.,

Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,  
 Plopper, F., Poldinger, A., Popovic, B., Primus, E., Pu, L.-L.,  
 Puzio, M., Quirós, J., Rachin, E., Reeves, K., Regier, M.A., Reigh, R.,  
 Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,  
 Rives, C., Rokey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,  
 Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,  
 Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D.,  
 Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,  
 Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,  
 Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umani, K.,  
 Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,  
 Wang, Q., Wang, S., Warren, R., Wei, X., White, F.,  
 Williams, G., Willson, R., Wleczyk, R., Wood, H., Worley, K.,  
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,  
 Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D. von  
 Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,  
 Weinstock, G. and Gibbs, R.A.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 253253)  
 Worley, K.C.  
 Direct Submission  
 Submitted (06-JUL-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 253253)  
 Rat Genome Sequencing Consortium.  
 Direct Submission  
 Submitted (13-MAY-2003) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On May 13, 2003 this sequence version replaced gi:24819062.  
 The sequence in this assembly is a combination of BAC based reads  
 and whole genome shotgun sequencing reads assembled using Atlas  
 (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described  
 in the feature table below represents a scaffold in the Atlas  
 assembly (a 'contig-scaffold'). Within each contig-scaffold,  
 individual sequence contigs are ordered and oriented, and separated  
 by sized gaps filled with Ns to the estimated size. The sequence  
 may extend beyond the ends of the clone and there may be sequence  
 contigs within a contig-scaffold that consist entirely of whole  
 genome shotgun sequence reads. Both end sequences and whole genome  
 shotgun sequence only contigs will be indicated in the feature  
 table.  
 ----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: http://www.hgsc.bcm.tmc.edu/  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: GMEX  
 Center clone name: CH230-157M5  
 ----- Summary Statistics  
 Assembly program: Atlas 3.0;  
 Consensus quality: 223138 bases at least Q40  
 Consensus quality: 225720 bases at least Q30  
 Consensus quality: 226900 bases at least Q20  
 Estimated insert size: 230245; sum-of-contigs estimation  
 Quality coverage: 7x in Q20 bases; sum-of-contigs estimation  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 3 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1 250704: contig of 250704 bp in length  
 \* 250705 250804: gap of unknown length  
 \* 250805 251997: contig of 1193 bp in length

\* 251998 252097: gap of unknown length  
 \* 252098 253253: contig of 1156 bp in length.  
 FEATURES  
 source  
 1..253253  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10116"  
 /clone="CH230-157M5"  
 misc\_feature  
 1..1369  
 /note="wgs end\_extension  
 clone\_end:T7"  
 331..4610  
 /note="wgs end\_extension  
 clone\_end:T7"  
 9695..10077  
 /note="clone\_boundary  
 clone\_end:T7"  
 misc\_feature  
 end\_sequence:BH286818"  
 complement(243837..244333)  
 /note="clone\_boundary  
 clone\_end:Sp6"  
 site:EcoRI  
 end\_sequence:BH286821"  
 246750..248811  
 /note="wgs end\_extension  
 clone\_end:Sp6"  
 248862..250704  
 /note="wgs end\_extension  
 clone\_end:Sp6"  
 BASE COUNT 64432 a 49523 c 48940 g 64796 t 25562 others  
 ORIGIN  
 Query Match 15.7%; Score 40.6; DB 2; Length 253253;  
 Best Local Similarity 52.0%; Pred. No. 0.46;  
 Matches 91; Conservative 0; Mismatches 84; Indels 0; Gaps 0;  
 QY 50 AAAAAAAGCTTACACCTTAACTCAACTCCAAACCCCTTCAGACAACTCGTTTCTT 109  
 Db 58024 ACATAATAGATACAGAAATCAATAATAGAACCAAAATGATTATCAACATTATCTTT 58083  
 QY 110 GAAACAAATCTTAACAGATAACTGCTGAGAAATACCGTTCCGAGCCCAACCCG 169  
 Db 58084 TAAAGGAATTTCAAACCTGGGTGCTCAGTATTATCTTTCAGTCCCACTCC 58143  
 QY 170 TACTGCAACTTTTATTGTGAACCTCCCAATTATGAGAAATCCCTTTTCCTCT 224  
 Db 58144 CTCTCCAACTCCTCTTGTGTCTCCCTCTCAAAACCAAGCCCTTTTCATCTAT 58198  
 Search completed: November 15, 2003, 03:34:12  
 Job time : 1150.51 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 14, 2003, 23:55:19 ; Search time 167.159 Seconds  
(without alignments)  
4182.570 Million cell updates/sec

Title: US-09-928-457-92

Perfect score: 259

Sequence: 1 AATTGCTATATCAAGTAGG.....TACTTACTGCCGCAAAATT 259

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq 19Jun03,\*

```
1: /SIDSL1/gcgdata/geneseq/geneseq-nuc/NA1980.DAT.*
2: /SIDSL1/gcgdata/geneseq/geneseq-nuc/NA1981.DAT.*
3: /SIDSL1/gcgdata/geneseq/geneseq-nuc/NA1982.DAT.*
4: /SIDSL1/gcgdata/geneseq/geneseq-nuc/NA1983.DAT.*
5: /SIDSL1/gcgdata/geneseq/geneseq-nuc/NA1984.DAT.*
6: /SIDSL1/gcgdata/geneseq/geneseq-nuc/NA1985.DAT.*
7: /SIDSL1/gcgdata/geneseq/geneseq-nuc/NA1986.DAT.*
8: /SIDSL1/gcgdata/geneseq/geneseq-nuc/NA1987.DAT.*
9: /SIDSL1/gcgdata/geneseq/geneseq-nuc/NA1988.DAT.*
10: /SIDSL1/gcgdata/geneseq/geneseq-nuc/NA1989.DAT.*
11: /SIDSL1/gcgdata/geneseq/geneseq-nuc/NA1990.DAT.*
12: /SIDSL1/gcgdata/geneseq/geneseq-nuc/NA1991.DAT.*
13: /SIDSL1/gcgdata/geneseq/geneseq-nuc/NA1992.DAT.*
14: /SIDSL1/gcgdata/geneseq/geneseq-nuc/NA1993.DAT.*
15: /SIDSL1/gcgdata/geneseq/geneseq-nuc/NA1994.DAT.*
16: /SIDSL1/gcgdata/geneseq/geneseq-nuc/NA1995.DAT.*
17: /SIDSL1/gcgdata/geneseq/geneseq-nuc/NA1996.DAT.*
18: /SIDSL1/gcgdata/geneseq/geneseq-nuc/NA1997.DAT.*
19: /SIDSL1/gcgdata/geneseq/geneseq-nuc/NA1998.DAT.*
20: /SIDSL1/gcgdata/geneseq/geneseq-nuc/NA1999.DAT.*
21: /SIDSL1/gcgdata/geneseq/geneseq-nuc/NA2000.DAT.*
22: /SIDSL1/gcgdata/geneseq/geneseq-nuc/NA2001A.DAT.*
23: /SIDSL1/gcgdata/geneseq/geneseq-nuc/NA2001B.DAT.*
24: /SIDSL1/gcgdata/geneseq/geneseq-nuc/NA2002.DAT.*
25: /SIDSL1/gcgdata/geneseq/geneseq-nuc/NA2003.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	259	100.0	259	AAV03600	Neisseria meningitidis
2	77.6	30.0	135	ABZ41613	N. gonorrhoeae nuc
3	56.8	21.9	1275	AAZ12106	Neisseria meningitidis
4	52.6	20.3	1572	AAAI5298	DNA encoding a pol
5	52.6	20.3	1575	AAZ12107	Neisseria gonorrhoeae
6	52.6	20.3	1575	AAAI5323	DNA encoding a pol
7	50.4	19.5	12438	AAAI5500	N. meningitidis pa
8	50.4	19.5	172325	AAZ1613	Neisseria meningitidis

9	50.4	19.5	349980	21	AAF21612	Neisseria meningitidis
10	50.4	19.5	837096	21	AAAI489	N. meningitidis pa
11	41.2	15.3	13326	24	ABL33713	Human immune system
12	38.2	14.7	15282	24	ABU70189	Chemically treated
13	38.2	14.7	15282	24	AAI61139	Human gene regulat
14	38	14.7	6944	24	ABK34027	Human DNA for stag
15	38	14.7	16509	24	ABL33320	Human immune system
16	37.8	14.6	5552	24	ABL33259	Human immune system
17	37.6	14.5	266	21	AAAI357	N. meningitidis Me
18	37.6	14.5	824	20	AAZ12104	Neisseria meningitidis
19	37.6	14.5	1575	20	AAZ12105	Neisseria meningitidis
20	36.4	14.1	8576	24	ABL34228	Human immune system
21	36.4	14.1	2200	24	ABL34190	Human immune system
22	36.4	14.1	6239	24	ABL70584	Chemically treated
23	36.4	14.1	6239	24	AAI61287	Human gene regulat
24	36.4	14.1	6239	24	ABK31387	Signal transductio
25	36.4	14.1	7316	24	ABN80241	Human chemically m
26	36.2	14.0	8781	24	ABL33687	Human immune system
27	36	13.9	335	22	AAI80630	Human polynucleoti
28	36	13.9	12133	24	ABN80323	Human chemically m
29	35.8	13.8	6099	24	ABL70398	Chemically treated
30	35.8	13.8	6099	24	AAI61346	Human gene regulat
31	35.8	13.8	6099	24	ABK31433	Signal transductio
32	35.8	13.8	7558	24	ABL33103	Human immune system
33	35.6	13.7	3499	21	AAI62298	Caenorhabditis ele
34	35.6	13.7	11996	24	ABL34493	Human metastasis a
35	35.4	13.7	621	24	ABQ46216	Oligonucleotide fo
36	35.4	13.7	621	24	ABQ46217	Oligonucleotide fo
37	35.4	13.7	6534	24	ABL32471	Human immune system
38	35.4	13.7	6066	24	ABN80011	Human chemically m
39	35.4	13.7	10569	24	ABN80264	Human gene regulat
40	35.4	13.7	10647	24	AAI61397	Tumour suppressor
41	35.2	13.6	9007	22	AAI46616	Oligonucleotide fo
42	34.8	13.4	623	24	ABQ43170	Oligonucleotide fo
43	34.8	13.4	623	24	ABQ43171	Oligonucleotide fo
44	34.8	13.4	625	24	ABQ40448	Oligonucleotide fo
45	34.8	13.4	625	24	ABQ40449	Oligonucleotide fo

#### ALIGNMENTS

RESULT 1  
AAV03600  
ID AAV03600 standard; DNA; 259 BP.  
XX  
AC AAV03600;  
XX  
DT 22-OCT-1998 (first entry)  
XX  
DE Neisseria meningitidis DNA sequence E45.  
XX  
KW N. gonorrhoeae; N. lactamica; chromosome 22491; region 1; region 2;  
KW region 3; pathogenicity; blood-brain barrier; diagnosis; infection;  
KW meningitis; ss.  
XX  
OS Neisseria meningitidis.  
XX  
PN WO9802547-A2.  
XX  
PD 22-JAN-1998.  
XX  
PF 11-JUL-1997; 97WO-FR01295.  
XX  
PR 12-JUL-1996; 96FR-0008768.  
XX  
PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.  
PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.  
PA (SMWK) SMITHKLINE BEECHAM.  
XX  
PI Nassif X, Tinsley C, Achtman M, Merker P, Ruelle J;  
PI Vinals C;  
XX

DR WPI; 1998-110594/10.  
 XX Genes present in *Neisseria meningitidis* but not other *Neisseria*  
 PT species - and related host cells, RNA, anti-sense sequences,  
 PT polypeptide(s) and antibodies, useful for diagnosing *Neisseria*  
 PT meningitidis infection and in protective vaccines  
 XX  
 PS Example 4; Pages 126-127; 150pp; French.  
 XX  
 CC AAV03575-606 represent sequences that are present in *Neisseria*  
 CC meningitidis and *N. gonorrhoeae* but not in *N. lactamica*, except for the  
 CC genes involved in biosynthesis of the capsule polysaccharide, fpaA or C,  
 CC opc, porA, rotamase, sequence IC1106, IgA protease, pillin, pilC,  
 CC proteins which bind transferrin and opacity proteins. The DNA sequences  
 CC are responsible for the differences in pathogenicity between *N.*  
 CC meningitidis and *N. gonorrhoeae*, specifically they include the genes that  
 CC allow *N. meningitidis* to cross the blood-brain barrier. DNA sequences  
 CC common to *N. meningitidis* and *N. gonorrhoeae*, but absent from *N.*  
 CC *lactamica*, are responsible for colonisation and penetration of the  
 CC mucosa. The DNA sequences can be used to produce probes and primers, and  
 CC antibodies produced against the encoded proteins are used in standard  
 CC hybridisation/immunoassay processes for diagnosis of *N. meningitidis*  
 CC infection, particularly meningitis.  
 XX  
 SQ Sequence 259 BP; 81 A; 71 C; 34 G; 73 T; 0 other;  
 Query Match 100.0%; Score 259; DB 19; Length 259;  
 Best Local Similarity 100.0%; Pred. No. 1.9e-71;  
 Matches 259; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 AATTGTATATCAAGTAGGATGGGCAATTTATGCTGACCTACAAAACCAAAACCAACCT 60  
 Db 1 AATTGTATATCAAGTAGGATGGGCAATTTATGCTGACCTACAAAACCAAAACCAACCT 60  
 QY 61 ACCACCTTAATCAACTCCACAAACCTCTTCAGACAACTCGTTTTTGAAGAAACATC 120  
 Db 61 ACCACCTTAATCAACTCCACAAACCTCTTCAGACAACTCGTTTTTGAAGAAACATC 120  
 QY 121 TGTAACAGATAACTGCTGAAGATAACCGTTGCGAGCCCCCAAAACCGTACTGCAACTT 180  
 Db 121 TGTAACAGATAACTGCTGAAGATAACCGTTGCGAGCCCCCAAAACCGTACTGCAACTT 180  
 QY 181 TTATGTGAACCTCCCAATATGAGAAATCCCTTTTCGTCCTCTTCTGATTGCGCCCT 240  
 Db 181 TTATGTGAACCTCCCAATATGAGAAATCCCTTTTCGTCCTCTTCTGATTGCGCCCT 240  
 QY 241 ACTTACTGCCAGCAAAATT 259  
 Db 241 ACTTACTGCCAGCAAAATT 259  
 RESULT 2  
 ABZ41613/c  
 ID ABZ41613 standard; DNA; 135 BP.  
 XX  
 AC ABZ41613;  
 XX  
 DT 07-MAR-2003 (first entry)  
 XX  
 DE *N. gonorrhoeae* nucleotide sequence SEQ ID 7815.  
 XX  
 KW Antibacterial; infection; vaccines; gene therapy; gene; ds.  
 XX  
 OS *Neisseria gonorrhoeae*.  
 XX  
 FN WO200079243-A2.  
 XX  
 FD 10-OCT-2002.  
 XX  
 PF 12-FEB-2002; 2002WO-IB02069.  
 XX  
 PR 12-FEB-2001; 2001GB-0003424.  
 XX

PA (CHIR-) CHIRON SPA.  
 XX  
 PI Fontana MR, Pizza M, Masignani V, Monaci E;  
 XX  
 DR WPI; 2003-058415/05.  
 DR P-PSDB; ABP0643.  
 XX  
 PT New protein from *Neisseria gonorrhoeae*, useful for the manufacture of a  
 PT medicament for treating or preventing *N. gonorrhoeae* infection  
 XX  
 PS Disclosure; Page 758; 815pp; English.  
 XX  
 CC The present invention relates to proteins from *Neisseria gonorrhoeae*.  
 CC Also disclosed are the nucleic acid molecules encoding the proteins and  
 CC antibodies that specifically bind to the proteins. The composition  
 CC comprising the protein, nucleic acid or antibody is useful for the  
 CC manufacture of a medicament for treating or preventing *N. gonorrhoeae*  
 CC infection, this may be in the form of a vaccine or gene therapy.  
 CC Sequences given in records ABZ37706-ABZ42016 represent nucleic acid  
 CC molecules of the invention.  
 XX  
 SQ Sequence 135 BP; 26 A; 21 C; 38 G; 50 T; 0 other;  
 Query Match 30.0%; Score 77.6; DB 25; Length 135;  
 Best Local Similarity 95.2%; Pred. No. 1.4e-14;  
 Matches 80; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 80 ACAAAACCTCTTCAGACAACTCGTTTTTGAAGAAACATCTGTAACAGATAACTGCTG 139  
 Db 84 ACAAAACCTCTTCAGACAACTCGTTTTTGAAGAAACATCTGTAACAGATAACTGCTG 25  
 QY 140 AAGAATACCGTTGCGAGCCCCAA 163  
 Db 24 AAGAATACCGTTGCGAGCCCCAA 1  
 RESULT 3  
 AAZ12106  
 ID AAZ12106 standard; DNA; 1275 BP.  
 XX  
 AC AAZ12106;  
 XX  
 DT 08-OCT-1999 (first entry)  
 XX  
 DE *Neisseria meningitidis* strain A complete ORF81 sequence.  
 XX  
 KW *Neisseria meningitidis*; *Neisseria gonorrhoeae*; antigen; vaccine;  
 KW treatment; *Neisseria* infection; meningitis; septicaemia; gonorrhea; ss.  
 XX  
 OS *Neisseria meningitidis*.  
 XX  
 FN WO9924578-A2.  
 XX  
 PD 20-MAY-1999.  
 XX  
 PF 09-OCT-1998; 98WO-IB01665.  
 XX  
 PR 01-SEP-1998; 98GB-0019016.  
 PR 06-NOV-1997; 97GB-0023516.  
 PR 14-NOV-1997; 97GB-0024190.  
 PR 18-NOV-1997; 97GB-0024386.  
 PR 27-NOV-1997; 97GB-0025158.  
 PR 10-DEC-1997; 97GB-0026147.  
 PR 14-JAN-1998; 98GB-0000759.  
 XX  
 PA (CHIR-) CHIRON SPA.  
 XX  
 PI Grandi G, Masignani V, Pizza M, Rappuoli R, Scarlato V;  
 XX  
 DR WPI; 1999-327407/27.  
 DR P-PSDB; AAY38652.  
 XX  
 PT Proteins from *Neisseria meningitidis* and *N. gonorrhoeae* useful for



PT diagnosis, treatment and prevention of infection  
 PS Claim 9; Page 206; 524pp; English.  
 XX  
 CC Nucleotide sequences AAZ11972-Z12358 represent open reading frames  
 CC (ORFs) of *Neisseria meningitidis* and *N. gonorrhoeae* which encode  
 CC antigenic proteins (see AAY38499-Y38944). The antigenic proteins, their  
 CC fragments, their nucleic acids and antibodies are used for diagnosis,  
 CC prevention (as vaccines) or treatment of *Neisseria* infections,  
 CC such as meningitis, septicaemia and gonorrhea. Both organisms  
 CC are closely related. Fragments of the nucleic acids are useful  
 CC as hybridisation probes and antisense reagents.  
 XX  
 SQ Sequence 1275 BP; 300 A; 304 C; 319 G; 352 T; 0 other;  
 Query Match 21.9%; Score 56.8; DB 20; Length 1275;  
 Best Local Similarity 96.7%; Pred. No. 1e-07;  
 Matches 58; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 200 ATGAGAAAATCCCTTTTCGTCCTCTTTCTGTTATTCGTCCTACTTACTGCCAGCGAAAT 259  
 Db 1 ATGAAAAAATCCCTTTTCGTCCTCTTTCTGTTATTCGTCCTACTTACTGCCAGCGAAAT 60  
 RESULT 4  
 AAA15298  
 ID AAA15298 standard; DNA; 1572 BP.  
 XX  
 AC AAA15298;  
 XX  
 DT 04-SEP-2000 (first entry)  
 XX  
 DE DNA encoding a polypeptide of a *Neisseria* pathogenic strain.  
 XX  
 KW Pathogenic strain; *Neisseria*; vaccine; *Neisseria* infection; ss.  
 XX  
 OS *Neisseria meningitidis*.  
 XX  
 FH Key Location/Qualifiers  
 FT 1..1572  
 FT CDS /\*tag= a  
 XX  
 PN WO200026375-A2.  
 XX  
 PD 11-MAY-2000.  
 XX  
 PF 28-OCT-1999; 99WO-FR02643.  
 XX  
 PR 30-OCT-1998; 98FR-0013693.  
 XX  
 XX (INMR ) PASTEUR MERIEUX SERUMS & VACCINS SA.  
 PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.  
 XX  
 PI Aujame L, Bouchardon A, Renaud-Mongenie G, Rokbi B, Nassif X;  
 PI Tinsley C, Perrin A;  
 XX  
 XX WPI; 2000-365622/31.  
 DR P-PSDB; AAY93267.  
 XX  
 PT New polypeptide specific for pathogenic *Neisseria* useful in therapeutic  
 PT or preventative vaccines and for diagnosis  
 XX  
 PS Claim 2; Page 56-59; 187pp; French.  
 XX  
 CC The present sequence encodes a protein that is specific for pathogenic  
 CC strains of *Neisseria*. The polynucleotides, polypeptides, or their  
 CC antigenic fragments, are used in vaccines to treat or protect against  
 CC *Neisseria* infections, particularly by *N. meningitidis*. The  
 CC polynucleotide sequence is also used for recombinant production of  
 CC the polypeptide and to produce attenuated *Neisseria* strains that  
 CC over-express it, or express it in a non-toxic mutant form.  
 XX  
 SQ Sequence 1572 BP; 380 A; 383 C; 395 G; 414 T; 0 other;

Query Match 20.3%; Score 52.6; DB 21; Length 1572;  
 Best Local Similarity 93.2%; Pred. No. 2.3e-06;  
 Matches 55; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 200 ATGAGAAAATCCCTTTTCGTCCTCTTTCTGTTATTCGTCCTACTTACTGCCAGCGAAAT 258  
 Db 1 ATGAAAAAATCCCTTTTCGTCCTCTTTCTGTTATTCGTCCTACTTACTGCCAGCGAAAT 59  
 RESULT 5  
 AAZ12107  
 ID AAZ12107 standard; DNA; 1575 BP.  
 XX  
 AC AAZ12107;  
 XX  
 DT 08-OCT-1999 (first entry)  
 XX  
 DE *Neisseria gonorrhoeae* complete ORF81 sequence.  
 XX  
 KW *Neisseria meningitidis*; *Neisseria gonorrhoeae*; antigen; vaccine;  
 KW treatment; *Neisseria* infection; meningitis; septicaemia; gonorrhea; ss.  
 XX  
 OS *Neisseria gonorrhoeae*.  
 XX  
 PN WO9924578-A2.  
 XX  
 PD 20-MAY-1999.  
 XX  
 PF 09-OCT-1998; 98WO-IB01665.  
 XX  
 PR 01-SEP-1998; 98GE-0019016.  
 PR 06-NOV-1997; 97GB-0023516.  
 PR 14-NOV-1997; 97GB-0024190.  
 PR 18-NOV-1997; 97GB-0024386.  
 PR 27-NOV-1997; 97GB-0025158.  
 PR 10-DEC-1997; 97GB-0026147.  
 PR 14-JAN-1998; 98GB-0000759.  
 XX  
 PA (CHIR-) CHIRON SPA.  
 XX  
 PI Grandi G, Masignani V, Pizza M, Rappuoli R, Scarlato V;  
 XX  
 DR WPI; 1999-327407/27.  
 DR P-PSDB; AAY38653.  
 XX  
 PT Proteins from *Neisseria meningitidis* and *N. gonorrhoeae* useful for  
 PT diagnosis, treatment and prevention of infection  
 XX  
 PS Claim 9; Page 208; 524pp; English.  
 XX  
 CC Nucleotide sequences AAZ11972-Z12358 represent open reading frames  
 CC (ORFs) of *Neisseria meningitidis* and *N. gonorrhoeae* which encode  
 CC antigenic proteins (see AAY38499-Y38944). The antigenic proteins, their  
 CC fragments, their nucleic acids and antibodies are used for diagnosis,  
 CC prevention (as vaccines) or treatment of *Neisseria* infections,  
 CC such as meningitis, septicaemia and gonorrhea. Both organisms  
 CC are closely related. Fragments of the nucleic acids are useful  
 CC as hybridisation probes and antisense reagents.  
 XX  
 SQ Sequence 1575 BP; 385 A; 387 C; 392 G; 411 T; 0 other;  
 Query Match 20.3%; Score 52.6; DB 20; Length 1575;  
 Best Local Similarity 93.2%; Pred. No. 2.3e-06;  
 Matches 55; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 200 ATGAGAAAATCCCTTTTCGTCCTCTTTCTGTTATTCGTCCTACTTACTGCCAGCGAAAT 258  
 Db 1 ATGAAAAAATCCCTTTTCGTCCTCTTTCTGTTATTCGTCCTACTTACTGCCAGCGAAAT 59  
 RESULT 6  
 AAA15323

```

ID AAA15323 standard; DNA; 1575 BP.
XX
AC AAA15323;
XX
DT 04-SEP-2000 (first entry)
XX
DE DNA encoding a polypeptide of a Neisseria pathogenic strain.
XX
KW Pathogenic strain; Neisseria; vaccine; Neisseria infection; ss.
XX
OS Neisseria gonorrhoeae.
FH Key Location/Qualifiers
FT CDS 1..1575
FT /*tag= a
XX
PN WO200026375-A2.
XX
PD 11-MAY-2000.
XX
PF 28-OCT-1999; 99WO-FR02643.
XX
PR 30-OCT-1998; 98FR-0013693.
XX
PA (INRM ) PASTEUR MERIEUX SERUMS & VACCINS SA.
PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX
PI Aujame L, Bouchardon A, Renauld-Mongenie G, Rokbi B, Nassif X;
PI Tinsley C, Perrin A;
XX
DR WPI; 2000-365622/31.
DR P-PSDB; AAY93293.
XX
PT New polypeptide specific for pathogenic Neisseria useful in therapeutic
PT or preventative vaccines and for diagnosis
XX
PS Claim 4; Page 137-139; 187pp; French.
XX
CC The present sequence encodes a protein that is specific for pathogenic
CC strains of Neisseria. The polynucleotides, polypeptides, or their
CC antigenic fragments, are used in vaccines to treat or protect against
CC Neisseria infections, particularly by N. meningitidis. The
CC polynucleotide sequence is also used for recombinant production of
CC the polypeptide and to produce attenuated Neisseria strains that
CC overexpress it, or express it in a non-toxic mutant form.
XX
SQ Sequence 1575 BP; 385 A; 387 C; 392 G; 411 T; 0 other;

Query Match 20.3%; Score 52.6; DB 21; Length 1575;
Best Local Similarity 93.2%; Pred. No. 2.3e-06;
Matches 55; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 200 ATGAGAAATCCCTTTCGCTCTTCTGTTATTCGTCCTACTACTGCCAGCAAT 258
DB 1 ATGAGAAATCCCTTTCGCTCTTCTGTTATTCGTCCTACTACTGCCAGCAAT 59

RESULT 7
AAA81550
ID AAA81550 standard; DNA; 12438 BP.
XX
AC AAA81550;
XX
DT 04-DEC-2000 (first entry)
XX
DE N. meningitidis partial DNA sequence gnm_97 SEQ ID NO:97.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
KW antigen; vaccine; diagnosis; infection; antibacterial; identification;
KW Meningococcus B; MenB; ds.
XX
OS Neisseria meningitidis.
XX

```

---

```

PN WO200022430-A2.
XX
PD 20-APR-2000.
XX
PF 08-OCT-1999; 99WO-US23573.
XX
PR 09-OCT-1998; 98US-0103794.
PR 30-APR-1999; 99US-0132068.
XX
PA (CHIR ) CHIRON CORP.
XX
PI Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;
PI Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;
PI Rappuoli R, Fizza M;
XX
DR WPI; 2000-318079/27.
XX
PT Isolated nucleotide sequences of Neisseria meningitidis which can be
PT used in the diagnosis and treatment of N. meningitidis infection and
PT other Neisserial infections, for example, N.gonorrhoea
XX
PS Claim 7; Page 1489-1492; 1760pp; English.
XX
CC The present invention describes methods of obtaining immunogenic
CC proteins from Neisseria genomic sequences. AAA81453 to AAA82414
CC represent specifically claimed Neisseria meningitidis genomic DNA
CC sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
CC Neisseria DNA sequences and their corresponding proteins; AAA81254 to
CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
CC isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
CC AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF
CC sequences, which are all used in the exemplification of the present
CC invention. The nucleic acid sequences, protein sequences, and antibodies
CC against them, can be used in the manufacture of a composition. The
CC composition can be used as a medicament (or in the manufacture of a
CC medicament) for treating, preventing or diagnosing infection due to
CC Neisserial bacteria. For example, some of the identified proteins could
CC be components of vaccines against Meningococcus B; against all serotypes;
CC and/or against all pathogenic Neisseriae. Identification of sequences
CC from the bacterium will also facilitate production of biological probes,
CC particularly organism-specific probes. Attempts to make efficacious
CC Meningococcus B vaccines have failed mainly due to antigen tolerance.
CC Multivalent vaccines have also been tried but none have successfully
CC overcome antigenic variability. The provision of further, complete
CC sequences may provide an opportunity to identify secreted or surface
CC exposed proteins that may be presumed targets for the immune system and
CC which are not antigenically variable or at least more conserved than
CC other more variable regions.
XX
SQ Sequence 12438 BP; 3661 A; 3104 C; 2565 G; 3108 T; 0 other;

Query Match 19.5%; Score 50.4; DB 21; Length 12438;
Best Local Similarity 75.0%; Pred. No. 2.3e-05;
Matches 63; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 176 AACTTTTATGTGAACCTCCCATATGAGAAAACCCCTTTTCGTCCTTCTCTATTCG 235
DB 2873 AATTTTATGTGAATATTAATGATGAAAAATCTTCTTACGCTTGTCTATTCG 2932

QY 236 TCCCTACTTACTGCCAGCAAAAT 259
DB 2933 TCTTTACTTACGCCAGCAAAAT 2956

RESULT 8
AAF21613
ID AAF21613 standard; DNA; 172325 BP.
XX
AC AAF21613;
XX
DT 13-MAR-2001 (first entry)
XX
DE Neisseria meningitidis B nucleotide sequence SEQ ID NO:114.

```

```

XX Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;
KW diagnosis; antigen; detection; infection; gene therapy; antibacterial;
KW ds.
XX Neisseria meningitidis.
OS WO200065791-A1.
XX 09-NOV-2000.
XX 08-MAR-2000; 2000WO-US05928.
XX 30-APR-1999; 99US-0132068.
PR 08-OCT-1999; 99WO-US23573.
PR 28-FEB-2000; 2000GB-0004695.
XX (CHIR ) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX Pizza M, Hickey E, Peterson J, Tettelin H, Venter JC, Masignani V;
PI Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V, Rappuoli R;
PI Frazer CW, Grandi G;
XX WPI; 2000-647603/62.
XX Neisseria meningitidis B full length genome sequence and open reading
frames are used to detect, treat and prevent Neisserial infections -
XX Claim 7; Appendix A; 692pp; English.
XX The present invention describes the full length genome of
CC Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607
CC to AAF21613 represent fragments of the NMB genomic sequence, as the
CC sequence was too long to go in a record on its own it was split into 8
CC sequences which overlap each other at the beginning and end of each
CC the beginning of AAF21607, the last 49980 bp of AAF21544 is repeated at
CC the beginning of AAF21607, and so on). AAF21545 to AAF21588 encode the
CC Neisseria proteins given in AAB58550 to AAB58593, and AAF21589 to
CC AAF21606 represent PCR primers which are used in the exemplification of
CC the present invention. The NMB genome and fragments from it have
CC antibacterial activity, and can be used in vaccines and gene therapy.
CC Neisseria nucleic acids, proteins and/or antibodies which binds to the
CC proteins can be used in compositions for treating or preventing infection
CC due to Neisserial bacteria or as a diagnostic reagent for detecting the
CC presence of Neisserial bacteria or of antibodies raised to Neisserial
CC bacteria. Computers, computer memory, computer storage medium or computer
CC databases can be used in a search to identify open reading frames (ORFs)
CC or coding sequences within the NMB genome. The DNA sequences provide
CC further opportunities to find antigenic or immunogenic proteins which are
CC more effective in vaccines than the outer membrane proteins currently
CC used.
XX SQ Sequence 172325 BP; 43072 A; 47583 C; 41465 G; 40205 T; 0 other;
Query Match 19.5%; Score 50.4; DB 21; Length 172325;
Best Local Similarity 75.0%; Pred. No. 5.7e-05;
Matches 63; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
QY 176 AACCTTTTATGTGAACTCCCATATGAGAAAATCCCTTTTCGCTCTTCTGTATTCG 235
DB 23053 AATTTTATGTGAAATATTAATGATGAAAAATCTTTCCTACGCTGTCTGTATTCG 23112
QY 236 TCCCTACTACTGCCAGCAAAATT 259
DB 23113 TCTTACTACCGCAGCAAAATT 23136
RESULT 9
AAF21612
ID AAF21612 standard; DNA; 349980 BP.
XX

```

```

AC AAF21612;
XX 13-MAR-2001 (first entry)
XX Neisseria meningitidis B nucleotide sequence SEQ ID NO:113.
XX Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;
KW diagnosis; antigen; detection; infection; gene therapy; antibacterial;
KW ds.
XX Neisseria meningitidis.
OS WO200065791-A1.
XX 09-NOV-2000.
XX 08-MAR-2000; 2000WO-US05928.
XX 30-APR-1999; 99US-0132068.
PR 08-OCT-1999; 99WO-US23573.
PR 28-FEB-2000; 2000GB-0004695.
XX (CHIR ) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX Pizza M, Hickey E, Peterson J, Tettelin H, Venter JC, Masignani V;
PI Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V, Rappuoli R;
PI Frazer CW, Grandi G;
XX WPI; 2000-647603/62.
XX Neisseria meningitidis B full length genome sequence and open reading
frames are used to detect, treat and prevent Neisserial infections -
XX Claim 7; Appendix A; 692pp; English.
XX The present invention describes the full length genome of
CC Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607
CC to AAF21613 represent fragments of the NMB genomic sequence, as the
CC sequence was too long to go in a record on its own it was split into 8
CC sequences which overlap each other at the beginning and end of each
CC the beginning of AAF21607, the last 49980 bp of AAF21544 is repeated at
CC the beginning of AAF21607, and so on). AAF21545 to AAF21588 encode the
CC Neisseria proteins given in AAB58550 to AAB58593, and AAF21589 to
CC AAF21606 represent PCR primers which are used in the exemplification of
CC the present invention. The NMB genome and fragments from it have
CC antibacterial activity, and can be used in vaccines and gene therapy.
CC Neisseria nucleic acids, proteins and/or antibodies which binds to the
CC proteins can be used in compositions for treating or preventing infection
CC due to Neisserial bacteria or as a diagnostic reagent for detecting the
CC presence of Neisserial bacteria or of antibodies raised to Neisserial
CC bacteria. Computers, computer memory, computer storage medium or computer
CC databases can be used in a search to identify open reading frames (ORFs)
CC or coding sequences within the NMB genome. The DNA sequences provide
CC further opportunities to find antigenic or immunogenic proteins which are
CC more effective in vaccines than the outer membrane proteins currently
CC used.
XX SQ Sequence 349980 BP; 86473 A; 95646 C; 85908 G; 81953 T; 0 other;
Query Match 19.5%; Score 50.4; DB 21; Length 349980;
Best Local Similarity 75.0%; Pred. No. 7.2e-05;
Matches 63; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
QY 176 AACCTTTTATGTGAACTCCCATATGAGAAAATCCCTTTTCGCTCTTCTGTATTCG 235
DB 23053 AATTTTATGTGAAATATTAATGATGAAAAATCTTTCCTACGCTGTCTGTATTCG 323112
QY 236 TCCCTACTACTGCCAGCAAAATT 259
DB 323113 TCTTACTACCGCAGCAAAATT 323136

```





```

XX SQ Sequence 15282 BP; 3617 A; 581 C; 4050 G; 7034 T; 0 other;
Query Match 14.7%; Score 38.2; DB 24; Length 15282;
Best Local Similarity 49.3%; Pred. No. 0.17;
Matches 100; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

QY 40 CTCARAAACCAAAACAACTACACACCTTAATCACTCCACAAACCCCTTCAGACAAC 99
Dd ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
8052 CAACAAACAAACCTAAATAATCCCCACCTCTCAATTTTAAATCTTACTTCCAAAAC 7993
QY 100 CTCGTTTTTGAAGAAACAATCTCTAAACAGATAACTGCTGAAGAATAACGTTGCCGAGCC 159
Dd ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7992 AAATCCCTTTAAACACAAACAACTAACTTCTAACTAAAAAACTATTCTAAAAAAAC 7933
QY 160 CCAAAACCCGCTACTGCACTTTTATGTAACCTCCCATATGAGAAATCCCTTTTCGT 219
Dd ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7932 CGAACTAAATCTATAACCCCAACCTACACACCTCCATCTACAAAAATATCCACITTTAA 7873
QY 220 CCTCTTCTGTTATCGCCCTAC 242
Dd ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7872 ATACAAAAAATATTCTCCCTAC 7850

RESULT 14
ABK34027/c
ID ABK34027 standard; DNA; 6944 BP.
XX AC ABK34027;
XX DT 18-JUN-2002 (first entry)
XX DE Human DNA for staging of Astrocytomas, complement, #57.
XX KW Human; ds; astrocytoma; cytostatic; staging; cysteine methylation; CpG;
XX KW bisulphite; brain tissue; MALDI; ESI; electron spray mass spectrometry;
XX KW matrix assisted laser desorption/ionization mass spectrometry.
XX OS Homo sapiens.
XX WO200202808-A2.
XX 10-JAN-2002.
XX 02-JUL-2001; 2001WO-EP07538.
XX 30-JUN-2000; 2000DE-1032529.
XX 01-SEP-2000; 2000DE-1043826.
XX (EPIG-) EPIGENOMICS AG.
XX Olek A. Piepenbrock C. Berlin K;
XX WPI; 2002-171649/22.
XX Novel chemically modified genomic DNA sequences, useful in the
XX characterisation, classification, differentiation, grading, staging,
XX treatment and/or diagnosis of astrocytomas or predisposition to
XX astrocytomas.
XX Claim 1; SEQ ID NO 114; 37pp; English.
XX The invention relates to a nucleic acid comprising a sequence (I) of at
XX least 18 bases in length of a segment of chemically pre-treated genomic
XX DNA which has any one of the sequences of (ABK33919-ABK34032) or its
XX complement. Also included are an oligonucleotide or peptide nucleic
XX acid (or set thereof) of at least 9 nucleotides which hybridises to (I),
XX primers for (II), probes for detecting cytosine methylation or single-
XX nucleotide polymorphisms (SNP) in (I), an array of oligomers
XX or peptide nucleic acids for analysing diseases associated with the
XX methylation states of the CpG dinucleotides of (I). The array is useful
XX for determining genetic and/or epigenetic parameters, classification,
XX differentiation, grading, staging, treatment and/or diagnosis of

```

```

CC astrocytomas, or the predisposition to astrocytomas by analysing cytosine
CC methylations, involves obtaining a biological sample containing genomic
CC DNA, extracting the genomic DNA, converting cytosine bases which are
CC unmethylated at the 5-position, in the genomic DNA sample, to uracil or
CC another base which is dissimilar to cytosine in terms of hybridisation
CC behaviour, by chemical treatment and amplifying chemically pre-treated
CC genomic DNA fragments using the array and a polymerase, where the
CC amplicates carry a detectable label. The method further involves
CC identifying methylation status of one or more cytosine positions, and
CC analysing methylation status of the cytosine positions by reference to
CC one or more data sets. The genomic DNA is chemically treated by using a
CC bisulphite, hydrogen sulphite or disulphite. The amplification
CC step amplifies DNA which is of particular interest in astrocytoma or
CC brain tissue, based on the specific genomic methylation status of brain
CC tissues, as opposed to background DNA. The amplicates carry a
CC fluorescent label or radionuclide. Optionally, the labels of the
CC amplicates are detachable molecule fragments having a typical mass
CC which are detected in a mass spectrometer. The fragments of chemically
CC pre-treated genomic DNA to be amplified, have a single positive or
CC negative charge for a better detectability in the mass spectrometer.
CC Preferably, the amplicates or fragments of the amplicates are
CC detected by matrix assisted laser desorption/ionization mass spectrometry
CC (MALDI) or using electron spray mass spectrometry (ESI). The
CC present sequence is one of the chemically pre-treated reference DNA
CC samples of the invention.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 6944 BP; 1633 A; 173 C; 1692 G; 3446 T; 0 other;
Query Match 14.7%; Score 38; DB 24; Length 6944;
Best Local Similarity 54.2%; Pred. No. 0.15;
Matches 77; Conservative 0; Mismatches 65; Indels 0; Gaps 0;
QY 44 AAACCAAAACAACTACACACCTTAATCAACTCCACAAACCTCTTCAGACAACCTCG 103
Dd ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4871 AAATCCAAAAAAACGACCTCTATAAAAAAATAAAAAACCCGCCGTAAACAGCA 4812
QY 104 TTTTGTAAAAACAATCTGTAAACAGATAACTGCTGAAGAATAACGTTGCCGAGCCCAA 163
Dd ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4811 AAACCTCAAAACCCGCTCCGAAAAAATCTAACGAATCTCAACCCCTCTCTCCSCCCA 4752
QY 184 AACCGTACTCCAACTTTTATT 185
Dd ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4751 AACTCCCACTCTTAAAAATATT 4730

```

```

RESULT 15
ABL33320/c
ID ABL33320 standard; DNA; 16509 BP.
XX AC ABL33320;
XX DT 26-MAR-2002 (first entry)
XX DE Human immune system associated gene SEQ ID NO: 1293.
XX KW Human; immune system disease; cytosine methylation; antiasthmatic;
XX KW antiarteriosclerotic; antianaemic; cytostatic; neurotropic;
XX KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
XX KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
XX KW antineoplastic; cancer; eye disease; arteriosclerosis; anaemia;
XX KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
XX KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
XX KW gene; ds.
XX OS Homo sapiens.
XX WO200200928-A2.
XX 03-JAN-2002.

```

```
XX 02-JUL-2001; 2001WO-EP07537.
PF
XX
XX 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
XX (EPIC-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
PI
XX WPI; 2002-130909/17.
DR
XX
XX Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation -
XX
XX Claim 1; SEQ ID NO 1293; 32pp + Sequence Listing; German.
PS
XX
XX The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.
XX
SQ Sequence 16509 BP; 3851 A; 379 C; 4791 G; 7488 T; 0 other;
Query Match 14.7%; Score 38; DB 24; Length 16509;
Best Local Similarity 53.1%; Pred. No. 0.2; Mismatches 90; Indels 1; Gaps 1;
Matches 103; Conservative 0;
QY 41 TACAAACCAAAACAACTACCACTTAACTCACTCCAC-AAACCCCTCTTCAGACAAC 99
Db 11520 TAAATATACAACTAACTACCACTTAAATCGCCACAAACCCCTCTTATAACATA 11461
QY 100 CTCGTTTTTGNAAACATCTGTAACAGATACTGCTGAAGATACCGTTGCCGAGCC 159
Db 11460 CTAATCAACAAACAAACAAATTAATAAATAAATACTTTAAATTAACGAACCTAAAC 11401
QY 160 CCAAAACCCGCTACTGCACTTTTATGTGAACCTCCCATTTATGAGAAATCCCTTTTCGT 219
Db 11400 TAAATCCGTTTCAATCTTTTACTTAATATTCGACCTAAACTAATTTATCCCTAAACT 11341
QY 220 CCTCTTCTGTATT 233
Db 11340 TTTTCTTTTAAAT 11327
```

Search completed: November 15, 2003, 00:35:44  
Job time : 169.159 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 14, 2003, 23:58:39 ; Search time 1727.81 Seconds  
(without alignments)  
3643.257 Million cell updates/sec

Title: US-09-928-457-92

Perfect score: 259

Sequence: 1 AATGTGTATATCAAGTAGG.....TACTTACTGCCAGCGAAATT 259

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : EST:

1: em\_estba.\*

2: em\_estum.\*

3: em\_estin.\*

4: em\_estmu.\*

5: em\_estov.\*

6: em\_estpl.\*

7: em\_estro.\*

8: em\_hic.\*

9: gb\_est1.\*

10: gb\_est2.\*

11: gb\_hic.\*

12: gb\_est3.\*

13: gb\_est4.\*

14: gb\_est5.\*

15: em\_estfun.\*

16: em\_estom.\*

17: em\_gss\_hum.\*

18: em\_gss\_inv.\*

19: em\_gss\_pln.\*

20: em\_gss\_vrt.\*

21: em\_gss\_fun.\*

22: em\_gss\_mam.\*

23: em\_gss\_mus.\*

24: em\_gss\_pro.\*

25: em\_gss\_rtd.\*

26: em\_gss\_pbg.\*

27: em\_gss\_vrl.\*

28: gb\_gss1.\*

29: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	45.2	17.5	1101	29	CNS00LT2
C 2	43.2	16.7	1201	13	EX419736
C 3	42.8	16.5	1181	13	EX363791
C 4	42	15.2	1101	29	CNS00DT7

C	5	40.6	15.7	581	12	BM962443
	6	39.6	15.3	1181	13	EX438271
	7	39.4	15.2	1078	28	EX2012
C	8	39.2	15.1	939	29	CNS00CNG
	9	38.6	14.9	712	13	EX416727
C	10	38.4	14.8	1201	13	EX355642
	11	38.4	14.8	1201	13	EX399985
C	12	38.2	14.7	1091	13	EX424950
	13	38.2	14.7	1101	29	CNS0182P
C	14	38	14.7	1173	13	EX456623
	15	37.8	14.6	1124	13	EX338623
C	16	37.8	14.6	1124	13	EX436282
	17	37.6	14.5	948	9	AL583463
C	18	37.6	14.5	1201	13	EX464312
	19	37.4	14.4	1101	29	CNS0172T
C	20	37.2	14.4	1196	13	EX342034
	21	37.2	14.4	2279	11	AK086359
C	22	37	14.3	623	13	BO931601
	23	37	14.3	1201	13	EX428660
C	24	36.8	14.2	619	28	AZ282736
	25	36.8	14.2	926	29	CNS008LM
C	26	36.8	14.2	1186	13	EX403722
	27	36.8	14.2	1201	9	AL548228
C	28	36.8	14.2	1201	13	EX361080
	29	36.6	14.1	1146	29	CNS021G2
C	30	36.4	14.1	588	28	AZ455602
	31	36.4	14.1	1101	29	CNS014NP
C	32	36.4	14.1	1201	13	EX405691
	33	36.2	14.0	538	9	AW117910
C	34	36.2	14.0	850	29	AG126628
	35	36	13.9	652	28	AZ784334
C	36	36	13.9	859	29	CNS004YY
	37	36	13.9	1072	13	EX383295
C	38	35.8	13.8	1111	13	EX338659
	39	35.8	13.8	374	9	AW455109
C	40	35.8	13.8	1101	29	CNS0006J
	41	35.8	13.8	1113	13	EX463898
C	42	35.8	13.8	1180	13	EX463188
	43	35.8	13.8	1201	9	AL563696
C	44	35.8	13.8	1201	13	EX354226
	45	35.6	13.7	255	12	BM958531

## ALIGNMENTS

RESULT 1  
CNS00LT2/c  
LOCUS  
DEFINITION  
Drosophila melanogaster genome survey sequence TET3 end of BAC:  
BACR48P19 of RFL1-98 library from Drosophila melanogaster (fruit fly), Genomic survey sequence.  
ACCESSION  
AL078714  
VERSION  
AL078714.1 GI:5102004  
KEYWORDS  
GSS.  
SOURCE  
Drosophila melanogaster (fruit fly)  
ORGANISM  
Drosophila melanogaster  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE  
1 (bases 1 to 1101)  
Genoscope.  
Direct Submission  
Submitted (11-JUN-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)

COMMENT  
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osogawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of



[illegible]

RESULT 4  
CNS00DT7  
LOCUS  
DEFINITION  
1101 bp DNA linear GSS 04-JUN-1999  
Drosophila melanogaster genome survey sequence TET3 end of BAC #  
BACR28J21 of RPC1-98 library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.  
ACCESSION  
AL075293  
VERSION  
AL075293.1 GI:4948467  
KEYWORDS  
GSS.  
SOURCE  
Drosophila melanogaster (fruit fly)  
ORGANISM  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephyroidea; Prosopterygota; Diptera; Brachycera; Muscomorpha;  
1 (bases 1 to 1101)  
REFERENCE  
Genoscope.  
AUTHORS  
Direct Submission  
TITLE  
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
JP 191 91006 EVRY cedex - FRANCE (E-mail : [secref@genoscope.cns.fr](mailto:secref@genoscope.cns.fr)  
- Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr))  
JOURNAL

web : [www.genoscope.cns.fr/](http://www.genoscope.cns.fr/)  
determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osagawa and Aaron Mammossier in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp; the same strain used for the BDGP's P1 and GST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.mcg.buffalo.edu/drosophila\\_bac.htm](http://bacpac.mcg.buffalo.edu/drosophila_bac.htm).

```

Found at: /home/ncs/biobase/ncs/blast/blast_db/ncs.new.
FEATURES
  source
    1. 1101
      /organism="Drosophila melanogaster"
      /mol_type="genomic DNA"
      /db_xref="taxon:7227"
      /clone="BACR28J21"
      /clone_lib="RPC1-98"
      /notes="end : Ter3"
BASE COUNT      312 a      157 c      328 t      224 others
ORIGIN
Query Match      16.2%   Score 42;   DB 29;   Length 1101;
Best Local Similarity 26.4%   Pred. No. 0.72;
Matches 64; Conservative 69; Mismatches 109; Indels 0; Gaps 0;

```

	6	530	66	590	126
QY	TG	TCT	CC	TY	AC
QY	AT	MC	CT	AA	GA
QY	AG	BC	AC	AT	AA
QY	AG	AT	CT	TA	AC
QY	TG	TC	CC	CT	GC
QY	AT	MC	CT	TA	AA
QY	AG	BC	AC	AT	AA
QY	AG	AT	CT	TA	AC
QY	TG	TC	CC	CT	GC
QY	AT	MC	CT	TA	AA
QY	AG	BC	AC	AT	AA
QY	AG	AT	CT	TA	AC
QY	TG	TC	CC	CT	GC
QY	AT	MC	CT	TA	AA
QY	AG	BC	AC	AT	AA
QY	AG	AT	CT	TA	AC
QY	TG	TC	CC	CT	GC
QY	AT	MC	CT	TA	AA
QY	AG	BC	AC	AT	AA
QY	AG	AT	CT	TA	AC
QY	TG	TC	CC	CT	GC
QY	AT	MC	CT	TA	AA
QY	AG	BC	AC	AT	AA
QY	AG	AT	CT	TA	AC
QY	TG	TC	CC	CT	GC
QY	AT	MC	CT	TA	AA
QY	AG	BC	AC	AT	AA
QY	AG	AT	CT	TA	AC
QY	TG	TC	CC	CT	GC
QY	AT	MC	CT	TA	AA
QY	AG	BC	AC	AT	AA
QY	AG	AT	CT	TA	AC
QY	TG	TC	CC	CT	GC
QY	AT	MC	CT	TA	AA
QY	AG	BC	AC	AT	AA
QY	AG	AT	CT	TA	AC
QY	TG	TC	CC	CT	GC
QY	AT	MC	CT	TA	AA
QY	AG	BC	AC	AT	AA
QY	AG	AT	CT	TA	AC
QY	TG	TC	CC	CT	GC
QY	AT	MC	CT	TA	AA
QY	AG	BC	AC	AT	AA
QY	AG	AT	CT	TA	AC
QY	TG	TC	CC	CT	GC
QY	AT	MC	CT	TA	AA
QY	AG	BC	AC	AT	AA
QY	AG	AT	CT	TA	AC
QY	TG	TC	CC	CT	GC
QY	AT	MC	CT	TA	AA
QY	AG	BC	AC	AT	AA
QY	AG	AT	CT	TA	AC
QY	TG	TC	CC	CT	GC
QY	AT	MC	CT	TA	AA
QY	AG	BC	AC	AT	AA
QY	AG	AT	CT	TA	AC
QY	TG	TC	CC	CT	GC
QY	AT	MC	CT	TA	AA
QY	AG	BC	AC	AT	AA
QY	AG	AT	CT	TA	AC
QY	TG	TC	CC	CT	GC
QY	AT	MC	CT	TA	AA
QY	AG	BC	AC	AT	AA
QY	AG	AT	CT	TA	AC
QY	TG	TC	CC	CT	GC
QY	AT	MC	CT	TA	AA
QY	AG	BC	AC	AT	AA
QY	AG	AT	CT	TA	AC
QY	TG	TC	CC	CT	GC
QY	AT	MC	CT	TA	AA
QY	AG	BC	AC	AT	AA
QY	AG	AT	CT	TA	AC
QY	TG	TC	CC	CT	GC
QY	AT	MC	CT	TA	AA
QY	AG	BC	AC	AT	AA
QY	AG	AT	CT	TA	AC
QY	TG	TC	CC	CT	GC
QY	AT	MC	CT	TA	AA
QY	AG	BC	AC	AT	AA
QY	AG	AT	CT	TA	AC
QY	TG	TC			

Db	650	WYTTMTCTMTCCMAAWATATMTCTYTMTHYTCYCYCWYYYCTCTCYCTYYYYYTYTT	709
Qy	186	GTGAACCTTCCCATATATCAGAAATACCTTTTCGTCCTCTCTTCTGTATTCGTCCTACTTA	245
Db	710	YHTTCTTACCTTCTYTTATTATCAATCTCTTTYYNNNNNNCCYCYCYYYYYY	769
Qy	246	CT 247	
Db	770	XY 771	

RESULT 5  
BM962443/c BM962443 581 bp mRNA linear EST 18-MAR-2002  
LOCUS cHAGK2T Ascidian hemocytes cDNA library Ciona intestinalis cdNA,  
DEFINITION mRNA sequence.  
ACCESSION BM962443 GI:19545870  
VERSION BM962443  
KEYWORDS EST.  
SOURCE Ciona intestinalis  
ORGANISM Ciona intestinalis  
Eukaryota; Metazoa; Chordata; Urochordata; Ascidacea; Enterogona;  
Phlebobranchia; Clonidae; Ciona.  
REFERENCE 1 (bases 1 to 581)  
AUTHORS Saida,K., Terajima,D., Satake,M., Kawazoe,Y. and Kasuya,A.  
TITLE Sequence tags expressed in hemocytes of Ciona intestinalis  
JOURNAL Unpublished  
COMMENT Contact : Masanobu Satake

Contact: Masaharu Satake  
 Department of Molecular Immunology  
 Institute of Development, Aging and Cancer, Tohoku University  
 Sairyo-machi 4-1, Aoba-ku, Sendai 980-8575, Japan  
 Tel: 81-22-717-8477  
 Fax: 81-22-717-8482  
 Email: satake@idac.tohoku.ac.jp.

FEATURES	Location/Qualifiers
source	1. .581

```

source
...:strain="clona intestinalis"
/organism="HANA"
/doi_xref="taxon.7719"
/doi_xref="hemocytes"
/cell_type="hemocytes"
/dev_stage="adult"
/clone_lib="Ascidian hemocytes cDNA library"
/notes="Vector: pZLI, 201L, 7 others"
BASE COUNT 202 a 88 C 201

```

BASE COUNT	202 a	88 c	83 g	201 t	7 others
ORIGIN					
Query Match		15.7%	Score 40.6;	DB 12;	Length 581;
Best Local Similarity		56.3%;	Pred. No. 1.6;		
Matches	76;	Conservative	0;	Mismatches	59;
				Indels	0;
				Gaps	0;

Qy	88	TCCTCAGACAACTCGTTTTTGGAAAAACAATCTGTAAACAGATAACTGCTGAAGAATAC	147
Db	380	TCATAAAATAAACAGAGTTTTTGGAAAAATTAACCTTTAAATAGATAACCGGTACAAACAG	321
Qy	148	CGTTGCGAGCCGCCAAAAACCGTACTGCACTTTTATTGTGAACTTCCCATTTATGAGAAA	207
Db	320	TTTTGCTGAACCTTTAAAAAGATAGATTATTATGTTTTTAAAGTCATTTTATAGTAG	261
Qy	208	ATCCCTTTTCGTCT	222
Db	260	ATCTCACTGTCGT	246

RESULT	6
EX438271	
LOCUS	
DEFINITION	BX438271 Homo sapiens NEUROBLASTOMA cDNA clone
ACCESSION	C90DAC06XF02 5-PRIME, mRNA sequence.
VERSION	BX438271
KEYWORDS	BX438271.1 GI:30771648
SOURCE	EST.
	Homo sapiens (human)



and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

## FEATURES

source  
1..939  
/organism="Drosophila melanogaster"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7227"  
/clone="BACR28H16"  
/note="end = TET3"

BASE COUNT 71 a 349 c 104 g 180 t 235 others

## ORIGIN

Query Match 15.1%; Score 39.2; DB 29; Length 939;  
Best Local Similarity 11.0%; Pred. No. 4.2;  
Matches 22; Conservative 108; Mismatches 70; Indels 0; Gaps 0;

QY 58 CCTACACCCCTTAATCAACTCCAAACCCCTCTTCAGACAACTCGTTTTTGAAGAAACA 117  
DB 353 CCTCCMMVHTTYMCCCCCCCMCMCHMTCCMMMMMMMMMMMMMMMMMMMMMMMM 412  
QY 118 ATCTGTAACAGATACTCGTGAAGATAACCGTTGCCGAGCCCAACCGTACTGCAA 177  
DB 413 MMTMMHMMMMMTTMMHMTTMMTMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM 472  
QY 178 CTTTATTGTGAACCTCCATTATGAGAAATCCCTTTTCCTCTCTTCTGTTATTCGTC 237  
DB 473 MTMMMMMMMMMTTMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM 532  
QY 238 CCTACTTACTGCCAGCGAAA 257  
DB 533 MMMMMMMMMMMMMMMMMMMMMM 552

## RESULT 9

LOCUS BX416727/c  
DEFINITION BX416727 Homo sapiens NEUROBLASTOMA 712 bp mRNA linear EST 15-MAY-2003  
CSODA011Y114 5-PRIME, mRNA sequence.

ACCESSION BX416727  
VERSION BX416727.1 GI:30765629

KEYWORDS EST.  
SOURCE Homo sapiens (human)

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
Li, W.B., Gruber, C., Jessee, J., and Polayes, D.  
Full-length cDNA libraries and normalization

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: [secre@genoscope.cns.fr](mailto:secre@genoscope.cns.fr), Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
Library was constructed by Life Technologies, a division of  
Invitrogen. Contact : Feng Liang Email : [fliang@lifetech.com](mailto:fliang@lifetech.com) URL :  
<http://fulllength.invitrogen.com/> Invitrogen Corporation 1600  
Faraday Avenue Genoscope sequence ID : CSODA011BE07Q91.

## FEATURES

## source

1..712  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CSODA011Y114"  
/tissue\_type="NEUROBLASTOMA"  
/clone\_lib="Homo sapiens NEUROBLASTOMA"  
/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed  
with a NotI-oligo(dT) primer. Five prime end enriched,  
double-strand cDNA was digested with Not I and cloned into  
the Not I and EcoRV sites of the pCMVSPORT 6 vector.  
Library was not normalized."  
BASE COUNT 154 a 53 c 50 g 148 t 307 others

## ORIGIN

Query Match 14.9%; Score 38.6; DB 13; Length 712;  
Best Local Similarity 3.5%; Pred. No. 5.9;  
Matches 5; Conservative 96; Mismatches 40; Indels 0; Gaps 0;

QY 38 ACCTACAAACCAAAACAACTTACCACCTTAATCAACTCCACAAACCTCTTCAGACA 97  
DB 612 MM 553  
QY 98 ACCTCGTTTTTGAAGAAACAATCTGTAAACAGATACTCTGAAGAAATACCGTTGCCGAG 157  
DB 552 MM 493

QY 158 CCCCAAAACCCGTACTGCAAC 178  
DB 492 CMCMMMMMMMMMMMMMMMMMMMMMM 472

## RESULT 10

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

## CONTACT

## UNPUBLISHED

## COMMENT

## GENOSCOPE

## BP 191 91006 EVRY cedex - France

Email: [secre@genoscope.cns.fr](mailto:secre@genoscope.cns.fr), Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)

## Library was constructed by Life Technologies, a division of

Invitrogen. Contact : Feng Liang Email : [fliang@lifetech.com](mailto:fliang@lifetech.com) URL :<http://fulllength.invitrogen.com/> Invitrogen Corporation 1600

## Faraday Avenue Genoscope sequence ID : CSODI002AH08Q91.

## FEATURES

## source

## 1..1201

## /organism="Homo sapiens"

## /mol\_type="mRNA"

## /db\_xref="taxon:9606"

## /clone="CSODI002Y015"

## /tissue\_type="PLACENTA COT 25-NORMALIZED"

## /clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"

## /note="1st strand cDNA was primed with a NotI-oligo(dT)

## primer. Five prime end enriched, double-strand cDNA was

## digested with Not I and cloned into the Not I and EcoR V

## sites of the pCMVSPORT 6 vector. Library was normalized."

## BASE COUNT 408 a 117 c 179 g 272 t 225 others

## ORIGIN

## Query Match

## Best Local Similarity

## Matches

## 44

## 1200

## 104

## 1140

## 164

## 1080

## 1021



	Ephydroidea; Drosophilidae; Drosophila.	
REFERENCE	1 (bases 1 to 1101)	
AUTHORS	Genoscope.	
TITLE	Direct Submission	
JOURNAL	Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)	
COMMENT	Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelosBAC11.	
FEATURES	Location/Qualifiers	
source	1..1101	
	/organism="Drosophila melanogaster"	
	/mol_type="genomic DNA"	
	/db_xref="taxon:7227"	
	/clone_lib="BACN37D10"	
	/clone_lib="DrosBAC"	
	/plasmid="pBelosBAC11"	
	/note="end : SP6"	
BASE COUNT	274 a 268 c 128 g 73 t 358 others	
ORIGIN		
	Query Match 14.7%; Score 38.2; DB 29; Length 1101;	
	Best Local Similarity 14.2%; Pred.No. 8;	
	Matches 31; Conservative 116; Mismatches 72; Indels 0; Gaps 0;	
QY	26 CATTATGCTGCAGCTACAAAACCAAAACCACTTAACTCACAACCTCCACAAAC 85	
DB	1097 HHHWATHMAMHCHWMHYMMHYMHYWHCHYMAAHAYMHYHMTVAHMCHAHHTWT 1038	
QY	86 CTCCTTCAGAACACTCGTTTGTGAAGAAACAATCTGTAACAGATACTGCTGAAGAAT 145	
DB	1037 HTYWTWCYMYAWYCCHHWMMYAWWWMMYAWMMHWHMMCMCMCHWWTMMH 978	
QY	146 ACGTGTCGGAGCCCCAACCCTGCTGCACTTATTTCTGTAACCTCCCATTTATGAGA 205	
DB	977 MCVCYCCWCHWCHHNTATTHHHYHYHWTCHMTTWYHYTMWHHHHAAM 918	
QY	206 AAATCCCTTTTCGCTCTTCTGCTGATTCGCCCTACT 244	
DB	917 AARHMAATHWYHMAHHTTYTYYTYYYYYYYY 879	
RESULT 14		
BX456623/c		
LOCUS	1173 bp mRNA linear EST 22-MAY-2003	
DEFINITION	BX456623 Homo sapiens THYMUS Homo sapiens cDNA clone CSOCAP03YC21	
ACCESSION	BX456623	
VERSION	BX456623.1 GI:31028793	
KEYWORDS	EST.	
SOURCE	Homo sapiens (human)	
ORGANISM	Homo sapiens	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
REFERENCE	1 (bases 1 to 1173)	
AUTHORS	Li,W.B., Gruber,C., Jesse,J. and Polayes,D.	
TITLE	Full-length cDNA libraries and normalization	
JOURNAL	Unpublished	
COMMENT	Contact : Genoscope Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - France Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr Library was constructed by Life Technologies, a division of Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/ invitroge Corporation 1600 Faraday Avenue Genoscope sequence ID : CSOCAP03AB1QP1. Location/Qualifiers	
FEATURES		
	1..1201	
	/organism="Homo sapiens"	
	/mol_type="mRNA"	
	/db_xref="taxon:9606"	
	/clone_lib="CSODI061YN05"	
	/tissue_type="PLACENTA COT 25-NORMALIZED"	
	/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"	
	/note="First strand cDNA was primed with a NotI-oligo(dT)"	
	primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."	
BASE COUNT	349 a 112 c 182 g 385 t 145 others	
ORIGIN		
	Query Match 14.7%; Score 38; DB 13; Length 1173;	
	Best Local Similarity 34.1%; Pred.No. 9.1;	
	Matches 73; Conservative 43; Mismatches 98; Indels 0; Gaps 0;	
QY	44 AAAACCAAACCAACCTTAGCACCTTAATCACTCCACAAACCTCTTCAGACAACTCG 103	
DB	1021 AMMCWAAMAHAHAANAAMWYCAMACTAMACAMYCYMMWCAACWHTHAMATTTTT 962	
QY	104 TTTTGTAAAAACAATCTGTAACAGATACTGCTGAAGATAACCGTTCGAGCCCCAA 163	
DB	961 TTTTATTTTAAAAWTTTACAAAMAAAAAATWYTCAAATTTTTTATTTTAAACCYA 902	
QY	164 AACCGTACTGCACTTTTATTGTAACCTTCCCATTATGAGAAATCCCTTTTGGTCTC 223	
DB	901 AAWAAAMWCHCACCCWAAAAWAMTTTTTHYTTTTTAMNMMAAATTTCMAAAACMWAAMC 842	
QY	224 TTTCTGTATTCGTCCTACTTACTGCGAGCGAAA 257	
DB	841 CTITTTTTTYCACCTTTTTTKCCYKARAAA 808	
RESULT 15		
BX338623		
LOCUS	1201 bp mRNA linear EST 02-MAY-2003	
DEFINITION	BX338623 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA clone CSODI061YN05 5-PRIME, mRNA sequence.	
ACCESSION	BX338623	
VERSION	BX338623.1 GI:30347019	
KEYWORDS	EST.	
SOURCE	Homo sapiens (human)	
ORGANISM	Homo sapiens	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
REFERENCE	1 (bases 1 to 1201)	
AUTHORS	Li,W.B., Gruber,C., Jesse,J. and Polayes,D.	
TITLE	Full-length cDNA libraries and normalization	
JOURNAL	Unpublished	
COMMENT	Contact : Genoscope Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - France Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 10133.f For more information about this cluster, see http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CSODI061CG03QP1&cluster=10133.f. Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/ invitroge Corporation 1600 Faraday Avenue Genoscope sequence ID : CSODI061CG03QP1. Location/Qualifiers	
FEATURES		
	1..1201	
	/organism="Homo sapiens"	
	/mol_type="mRNA"	
	/db_xref="taxon:9606"	
	/clone_lib="CSODI061YN05"	
	/tissue_type="PLACENTA COT 25-NORMALIZED"	
	/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"	
	/note="First strand cDNA was primed with a NotI-oligo(dT)"	
	primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."	
BASE COUNT	349 a 112 c 182 g 385 t 145 others	
ORIGIN		
	Query Match 14.7%; Score 38; DB 13; Length 1173;	
	Best Local Similarity 34.1%; Pred.No. 9.1;	
	Matches 73; Conservative 43; Mismatches 98; Indels 0; Gaps 0;	
QY	44 AAAACCAAACCAACCTTAGCACCTTAATCACTCCACAAACCTCTTCAGACAACTCG 103	
DB	1021 AMMCWAAMAHAHAANAAMWYCAMACTAMACAMYCYMMWCAACWHTHAMATTTTT 962	
QY	104 TTTTGTAAAAACAATCTGTAACAGATACTGCTGAAGATAACCGTTCGAGCCCCAA 163	
DB	961 TTTTATTTTAAAAWTTTACAAAMAAAAAATWYTCAAATTTTTTATTTTAAACCYA 902	
QY	164 AACCGTACTGCACTTTTATTGTAACCTTCCCATTATGAGAAATCCCTTTTGGTCTC 223	





```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on:      November 15, 2003, 00:06:39 ; Search time 40.3413 Seconds
              (without alignments)
              2833.774 Million cell updates/sec

Title:      US-09-928-457-92
Perfect score: 259
Sequence:    1 AATTGCTATATCAAGTAGG.....TACTTACTCCAGCGAATT 259

Scoring table:  IDENTITY NUC
                  Gapcp 10'-0 , Gapext 1.0

```

C	1	35.6	13.7	3499	3	US-08-85U-076-43	Sequence 43, Appl
	2	33	12.7	2229	4	US-09-45A-495-1	Sequence 1, Appl
	3	31	12.0	1553	4	US-09-461-325-74	Sequence 74, Appl
	4	30.4	11.7	148567	4	US-09-801-876B-3	Sequence 3, Appl
	5	29.6	11.4	5183	1	US-08-45S-568-3	Sequence 3, Appl
	6	29.6	11.4	5183	2	US-08-399-411-3	Sequence 3, Appl
	7	29.6	11.4	5868	3	US-08-516-859A-3	Sequence 3, Appl
	8	29.6	11.4	5868	4	US-09-586-472-3	Sequence 3, Appl
	9	29.6	11.4	5868	4	US-09-528-706-3	Sequence 3, Appl
	10	29.4	11.4	5532	4	US-08-851-557B-52	Sequence 52, Appl
	11	29.4	11.4	6836	4	US-08-976-259-73	Sequence 73, Appl
	12	29.4	11.4	7515	4	US-08-851-567B-11	Sequence 11, Appl
	13	29.2	11.3	750	4	US-09-328-352-1687	Sequence 1687, Ap
	14	29	11.2	2102	4	US-09-620-312D-487	Sequence 487, Appl
	15	29	11.2	10607	1	US-08-078-090-3	Sequence 3, Appl
	16	28.8	11.1	369	4	US-09-713-016-81	Sequence 81, Appl
	17	28.8	11.1	960	3	US-08-975-628-1	Sequence 1, Appl
	18	28.8	11.1	960	4	US-09-427-774-1	Sequence 1, Appl
	19	28.8	11.1	1113	4	US-09-134-001C-1556	Sequence 1556, Ap
	20	28.8	11.1	3001	4	US-09-539-333D-131	Sequence 131, App
	21	28.8	11.1	3001	4	US-09-539-333D-137	Sequence 137, Appl
	22	28.8	11.1	1664976	4	US-08-916-421B-1	Sequence 1, Appl
	23	28.4	11.0	3308	3	US-08-714-918-68	Sequence 68, Appl
	24	28.4	11.0	3308	3	US-09-265-315-68	Sequence 68, Appl
	25	28.4	11.0	3308	3	US-09-265-315-68	Sequence 68, Appl
	26	28.4	11.0	3308	3	US-09-266-417-68	Sequence 68, Appl
	27	28.4	11.0	1230025	4	US-09-198-452A-1	Sequence 1, Appl

RESULT 2  
US-09-454-495-1/c  
; Sequence 1, Application US/09454495  
; Patent No. 6576759  
; GENERAL INFORMATION:  
; APPLICANT: Ready, Gurucharan  
; APPLICANT: Zeng, Hong

```
Query Match      12.0%; Score 31; DB 4; Length 1553;
Best Local Similarity 53.8%; P-Val 1.5e-06;
Matches 64; Conservative 0; Mismatches 55; Indels 0; Gaps 0;
```

APPLICATION NUMBER: US 08/399,411

```

; FILING DATE: 06-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1264
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5183 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..5158
US-08-459-568-3

Query Match 11.4%; Score 29.6; DB 1; Length 5183;
Best Local Similarity 57.6%; Pred. No. 7;
Matches 53; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

Qy 100 CTCGTTTTTGAAGAACATCTGTAACAGATACTGCTGAAGATACCGTTGCCGAGCC 159
Db 5002 CTCCTTTGTTCTCACTCAAGTCAGCAGCAGCTGCCAGCTCGCGGTGACTGGCC 4943

Qy 160 CCAAAACCCGCTACTGCAACTTTTATTGTGAAC 191
Db 4942 CCCCACTCCGCTCTCTAGATTTTATTGAAC 4911

RESULT 6
US-08-399-411-3/c
; Sequence 3, Application US/08399411
; Patent No. 5831008
; GENERAL INFORMATION:
; APPLICANT: Huang, Shi
; TITLE OF INVENTION: Retinoblastoma Protein - Interacting
; NUMBER OF SEQUENCES: 93
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/399,411
; FILING DATE: 06-MAR-1995
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5183 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..5158

```

```

US-08-399-411-3

Query Match 11.4%; Score 29.6; DB 2; Length 5183;
Best Local Similarity 57.6%; Pred. No. 7;
Matches 53; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

Qy 100 CTCGTTTTTGAAGAACATCTGTAACAGATACTGCTGAAGATACCGTTGCCGAGCC 159
Db 5002 CTCCTTTGTTCTCACTCAAGTCAGCAGCAGCTGCCAGCTCGCGGTGACTGGCC 4943

Qy 160 CCAAAACCCGCTACTGCAACTTTTATTGTGAAC 191
Db 4942 CCCCACTCCGCTCTCTAGATTTTATTGAAC 4911

RESULT 7
US-08-516-859A-3/c
; Sequence 3, Application US/08516859A
; Patent No. 6069231
; GENERAL INFORMATION:
; APPLICANT: Huang, Shi
; TITLE OF INVENTION: Retinoblastoma Protein - Interacting
; NUMBER OF SEQUENCES: 106
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/516,859A
; FILING DATE: 18-AUG-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/399,411
; FILING DATE: 06-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/292,683
; FILING DATE: 18-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1776
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5868 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 121..5278
US-08-516-859A-3

Query Match 11.4%; Score 29.6; DB 3; Length 5868;
Best Local Similarity 57.6%; Pred. No. 7.4;
Matches 53; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

Qy 100 CTCGTTTTTGAAGAACATCTGTAACAGATACTGCTGAAGATACCGTTGCCGAGCC 159
Db 5122 CTCCTTTGTTCTCACTCAAGTCAGCAGCAGCTGCCAGCTCGCGGTGACTGGCC 5063

Qy 160 CCAAAACCCGCTACTGCAACTTTTATTGTGAAC 191

```



```
; APPLICANT: Merlo, Donald J.
; APPLICANT: Orr, Gregory L.
; APPLICANT: Roberts, Jean L.
; APPLICANT: Strickland, James A.
; APPLICANT: Guo, Lining
; APPLICANT: Ciche, Todd A.
; APPLICANT: Sukhapinda, Kitisi
; TITLE OF INVENTION: Insecticidal Protein Toxins From Photorhabdus
; NUMBER OF SEQUENCES: 88
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dow Agrosciences Patent Department
; STREET: 9330 Zionsville Road
; CITY: Indianapolis
; STATE: IN
; COUNTRY: US
; ZIP: 46268
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/851.567B
; FILING DATE: 05-MAY-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/063.615
; FILING DATE: 18-MAY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/395.497
; FILING DATE: 28-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/007.255
; FILING DATE: 06-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/608.423
; FILING DATE: 28-FEB-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/705.484
; FILING DATE: 28-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Seav, Nicholas J
; REGISTRATION NUMBER: 27386
; REFERENCE/DOCKET NUMBER: 960296.93804
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 608-251-5000
; TELEFAX: 608-251-9166
; INFORMATION FOR SEQ ID NO: 52:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5532 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-851-567B-52

Query Match 11.4%; Score 29.4; DB 4; Length 5532;
Best Local Similarity 58.6%; Pred. No. 8.4;
Matches 51; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

Qy 124 AAACAGATAACTGCTGAAGATACCGTTGCCGAGCCCAAAACCGTACTGCAACTTTTA 183
Db 2828 ACATAGTTTCTGGATATAGACAGTTCAGAGACACCAGCCCAAGTACTGTACGTTTA 2769

Qy 184 TTGTGAACCTCCCATTTAGAGAAATC 210
Db 2768 TTGTAACTGCTCCAGTCAGTGAAGAAC 2742

RESULT 11
US-08-976-259-73
; Sequence 73, Application US/08976259
; Patent No. 6316609
```

```
; GENERAL INFORMATION:
; APPLICANT: Dillon, Patrick J.
; APPLICANT: Choi, Gil H.
; APPLICANT: Welch, Rodney A.
; TITLE OF INVENTION: Nucleotide Sequence of Escherichia coli
; Patent No. 6316609
; NUMBER OF SEQUENCES: 142
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Ave, N.W., Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/976.259
; FILING DATE: Herewith
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/031.626 AND US 60/061.953
; ATTORNEY/AGENT INFORMATION:
; NAME: Steffe, Eric K.
; REGISTRATION NUMBER: 36.688
; REFERENCE/DOCKET NUMBER: 1488.0740002/EKS/CBM
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 73:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6836 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-976-259-73

Query Match 11.4%; Score 29.4; DB 4; Length 6836;
Best Local Similarity 56.8%; Pred. No. 9.1;
Matches 54; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

Qy 126 ACAGATAACTGCTGAAGATACCGTTGCCGAGCCCAAAACCGTACTGCAACTTTTAT 185
Db 292 ACTGATTGCAGTTCAGGATACCGTTTCCCAACACACACCTGTTCATCATCTTTTAA 351

Qy 186 GTGAACCTCCCATTTAGAGAAATCCCTTTTCGTC 220
Db 352 TGTGAGGCACAGTACGAAACAAACAAAGTTTCATC 386

RESULT 12
US-08-851-567B-11/c
; Sequence 11, Application US/08851567B
; Patent No. 6528484
; GENERAL INFORMATION:
; APPLICANT: Ensign, Gerald C
; APPLICANT: Bowen, David J
; APPLICANT: Petell, James
; APPLICANT: Fatig, Raymond
; APPLICANT: Schoonover, Sue
; APPLICANT: French-Constant, Richard
; APPLICANT: Rocheleau, Thomas A.
; APPLICANT: Blackburn, Michael B.
; APPLICANT: Hey, Timothy D.
; APPLICANT: Merlo, Donald J.
; APPLICANT: Orr, Gregory L.
; APPLICANT: Roberts, Jean L.
; APPLICANT: Strickland, James A.
; APPLICANT: Guo, Lining
; APPLICANT: Ciche, Todd A.
```

APPLICANT: Sukhapinda, Kitisri  
TITLE OF INVENTION: Insecticidal Protein Toxins From Photorhabdus  
NUMBER OF SEQUENCES: 88  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dow Agrosciences Patent Department  
STREET: 9330 Zionsville Road  
CITY: Indianapolis  
STATE: IN  
COUNTRY: US  
ZIP: 46268  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/851,567B  
FILING DATE: 05-MAY-1997  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/063,615  
FILING DATE: 18-MAY-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/395,497  
FILING DATE: 28-FEB-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/007,255  
FILING DATE: 06-NOV-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/608,423  
FILING DATE: 28-FEB-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/705,484  
FILING DATE: 28-AUG-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Seay, Nicholas J  
REGISTRATION NUMBER: 27386  
REFERENCE/DOCKET NUMBER: 960296.93804  
TELEPHONE: 608-251-5000  
TELEFAX: 608-251-9166  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7515 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..7515  
US-08-851-567B-11

Query Match 11.4%; Score 29.4; DB 4; Length 7515;  
Best Local Similarity 59.6%; Pred. No. 9.5;  
Matches 51; Conservative 0; Mismatches 36; Indels 0; Gaps 0;  
QY 124 AACAGATACTGCTGAAGATACCGTTCGCGAGCCCAAAACCCGCTACTGCAACTTTTA 183  
DB 3089 ACATAGTTTCTGATATAGACCACTTCAGAGACACACCCCAAGTACTGTAACTTTA 3030  
QY 184 TTGTGAACCTCCCATTTATGAGAAATC 210  
DB 3029 TTGTAACTCCCACTTCAGTGAAGAAC 3003

RESULT 13  
US-09-328-352-1687/c  
; Sequence 1687, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER

; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC99-03PA  
; CURRENT APPLICATION NUMBER: US/09/328,352  
; CURRENT FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 1687  
; LENGTH: 750  
; TYPE: DNA  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-1687

Query Match 11.3%; Score 29.2; DB 4; Length 750;  
Best Local Similarity 59.8%; Pred. No. 4.3;  
Matches 49; Conservative 0; Mismatches 33; Indels 0; Gaps 0;  
QY 82 AAACCTCTTCAGACAACTCGTTTGTGAAAAACAATCTGTAACACAGATAACTGCTGAA 141  
DB 350 AAACCTTGCCTTAACCAAGCCATTTCTCAAAATAATATTTAAGCCCATCAGTCTGCTAA 291  
QY 142 GAATACCGTTGCCGAGCCCA 163  
DB 290 GAAGCCCAACATAACACCCA 269

RESULT 14  
US-09-620-312D-487/c  
; Sequence 487, Application US/09620312D  
; Patent No. 6569662  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Liu, Chenghua  
; APPLICANT: Asundi, Vinod  
; APPLICANT: Zhang, Jie  
; APPLICANT: Ren, Feiyang  
; APPLICANT: Chen, Rui-hong  
; APPLICANT: Zhao, Qing A.  
; APPLICANT: Wehrman, Tom  
; APPLICANT: Xue, Aidong J.  
; APPLICANT: Yang, Yonghong  
; APPLICANT: Wang, Jian-Rui  
; APPLICANT: Zhou, Ping  
; APPLICANT: Ma, Yungqing  
; APPLICANT: Wang, Dunrui  
; APPLICANT: Wang, Zhiwei  
; APPLICANT: John, Tillinghast  
; APPLICANT: Drmanac, Radoje T.  
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and  
; FILE REFERENCE: Polypeptides  
; FILE REFERENCE: 784C1P2B  
; CURRENT FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: US/09/620,312D  
; PRIOR FILING DATE: 2000-04-25  
; PRIOR APPLICATION NUMBER: 09/552,317  
; PRIOR FILING DATE: 2000-01-21  
; NUMBER OF SEQ ID NOS: 1105  
; SOFTWARE: pt FL\_genes Version 1.0  
; SEQ ID NO 487  
; LENGTH: 2102  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; NAME/KEY: CDS  
; LOCATION: (309)..(1283)  
US-09-620-312D-487

Query Match 11.2%; Score 29; DB 4; Length 2102;  
Best Local Similarity 63.8%; Pred. No. 7.6;  
Matches 44; Conservative 0; Mismatches 25; Indels 0; Gaps 0;  
QY 158 CCCCACCAACCCGCTACTGCAACTTTTATGTGAACCTCCCATTTATGAGAAATCCCTTTC 217  
DB 461 CTCCAAGATGCTTCTGCACTTTTATATACATTCACCTGATTATAATTTCTTCTC 402

Qy 218 GTCCTCTTT 226  
Db 401 AGACTCTGT 393

RESULT 15  
US-08-078-090-3/c  
; Sequence 3, Application US/08078090  
; Patent No. 5739407  
; GENERAL INFORMATION:  
; APPLICANT: BERGSTROM, SVEN  
; APPLICANT: HERNELL, OLLE  
; APPLICANT: LOENNERDAL, BO  
; APPLICANT: HJALMARSSON, KARIN  
; APPLICANT: HANSSON, LENNART  
; APPLICANT: TOERNELL, JAN  
; APPLICANT: STROEMQUIST, MATS  
; TITLE OF INVENTION: HUMAN BETA-CASEIN PROCESS FOR PRODUCING  
; TITLE OF INVENTION: IT AND USE THEREOF  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROWDY AND NEIMARK  
; STREET: 419 SEVENTH STREET, N.W.  
; CITY: WASHINGTON  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/078,090  
; FILING DATE: 19930618  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: WO PCT/DK92/00236  
; FILING DATE: 19-AUG-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: WO PCT/DK91/00233  
; FILING DATE: 19-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: COOPER, IVER P.  
; REGISTRATION NUMBER: 28,005  
; REFERENCE/DOCKET NUMBER: BERGSTROM2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)628-5197  
; TELEFAX: (202)737-3528  
; TELEX: 248633  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 10607 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: join(4804..4854, 5720..5746, 6726..6746, 6845  
; LOCATION: ..6886, 7991..8521, 9440..9445)  
US-08-078-090-3

Query Match 11.2%; Score 29; DB 1; Length 10607;  
Best Local Similarity 53.0%; Pred. No.15;  
Matches 62; Conservative 0; Mismatches 55; Indels 0; Gaps 0;  
Qy 6 TGTATATCAAGTAGATGGGATTTATGCTTCACCTCAAAAACCAAAAACACCTACCAC 65  
Db 6674 TATAAATCAGGAATAAAGGCAATTAATCTTTGATATATCATCAATCACTTTGTATAT 6615  
Qy 66 CTTTAACTCACTCCACAAACCCCTCTTCAGACAACCTCGTTTTTTTGAACAAATCTG 122

Db 6614 CATTAACCTTATGTATTATTAACCTGTTCTATATGAGCTGGTTATGTAAAGAAAGTCAG 6558  
Search completed: November 15, 2003, 08:09:36  
Job time : 44.3413 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 15, 2003, 00:09:34 ; Search time 146.667 Seconds  
(without alignments)  
5771.357 Million cell updates/sec

Title: US-09-928-457-92

Perfect score: 259

Sequence: 1 AATTGTTATATCAAGTAGG.....TACTTACTGCCACGGAATT 259

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2169961 seqs, 1634102185 residues

Total number of hits satisfying chosen parameters: 4339922

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:
- 2: /cgn2\_6/ptodata/1/pubpna/PCCT\_NEW\_PUB.seq:
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:
- 6: /cgn2\_6/ptodata/1/pubpna/PCCTUS\_PUBCOMB.seq:
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:
- 9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq:
- 10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:
- 11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:
- 12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:
- 13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:
- 14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:
- 15: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:
- 16: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:
- 17: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	259	100.0	259	10	US-09-928-457-92
2	41.2	15.9	13326	12	US-10-311-455-1686
3	38	14.7	6944	14	US-10-172-086-112
4	38	14.7	16509	12	US-10-311-455-1293
5	37.8	14.6	5552	12	US-10-311-455-1232
6	37	14.3	3673778	12	US-10-312-841-1
7	36.6	14.1	8576	12	US-10-311-455-2201
8	36.4	14.1	2200	12	US-10-311-455-2163
9	36.2	14.0	8781	12	US-10-311-455-1660
10	35.8	13.8	7758	9	US-09-205-858-43
11	35.6	13.7	3499	9	US-09-844-353A-43
12	35.6	13.7	3499	12	US-09-963-693-43
13	35.6	13.7	11996	12	US-10-240-485-46
14	35.6	13.7	6534	12	US-10-311-455-444
15	35.4	13.7	5929	12	US-10-311-455-682
16	34.8	13.4	5929	12	US-10-311-455-682

C 17	34.8	13.4	7491	12	US-10-311-455-1557
C 18	34.8	13.4	9293	12	US-10-240-453-33
C 19	34.8	13.4	9293	14	US-10-239-676-25
C 20	34.8	13.4	3673778	12	US-10-312-841-2
C 21	34.4	13.3	6631	12	US-10-240-453-213
C 22	34.4	13.3	9084	12	US-10-311-455-1578
C 23	34.4	13.3	11131	12	US-10-204-708-28
C 24	34.4	13.3	11131	12	US-10-311-455-726
C 25	34.4	13.3	13084	12	US-10-311-455-2151
C 26	34.2	13.2	7934	12	US-10-240-453-169
C 27	34.2	13.2	7934	14	US-10-239-676-153
C 28	34.2	13.2	13125	12	US-10-311-455-1200
C 29	34.2	13.2	13125	12	US-10-240-485-110
C 30	34.2	13.2	14353	12	US-10-311-455-2407
C 31	34	13.1	6396	12	US-10-311-455-2220
C 32	34	13.1	17967	12	US-10-311-455-988
C 33	33.8	13.1	5678	12	US-10-311-455-1111
C 34	33.8	13.1	8301	12	US-10-311-455-25
C 35	33.8	13.1	12733	12	US-10-311-455-63
C 36	33.6	13.0	764	12	US-10-027-632-150062
C 37	33.6	13.0	764	12	US-10-027-632-150063
C 38	33.6	13.0	764	12	US-10-027-632-150064
C 39	33.6	13.0	764	12	US-10-027-632-150065
C 40	33.6	13.0	764	13	US-10-027-632-150062
C 41	33.6	13.0	764	13	US-10-027-632-150063
C 42	33.6	13.0	764	13	US-10-027-632-150064
C 43	33.6	13.0	764	13	US-10-027-632-150065
C 44	33.6	13.0	4993	12	US-10-311-455-2076
C 45	33.4	12.9	6801	12	US-10-204-708-61

ALIGNMENTS

RESULT 1

US-09-928-457-92

Sequence 92, Application US/09928457

Patent No. US2002016403A1

GENERAL INFORMATION:

APPLICANT:

TITLE OF INVENTION: DNA, specific proteins and peptides

TITLE OF INVENTION: of the Neisseria meningitidis species bacteria, method for obtaining them and their biological application.

TITLE OF INVENTION: for obtaining them and their biological application.

NUMBER OF SEQUENCES: 99

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30 (OEB)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/928,457

FILING DATE: 2001-08-14

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/214,759

FILING DATE: 199-12-10

INFORMATION FOR SEQ ID NO: 92:

SEQUENCE CHARACTERISTICS:

LENGTH: 259 base pairs

TYPE: nucleotide

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHEetical: NO

ANTI-SENSE: NO

US-09-928-457-92

Query Match 100.0%; Score 259; DB 10; Length 259;

Best Local Similarity 100.0%; Pred. No. 3.2e-70;

Matches 259; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AATTGTTATATCAAGTAGGATGGGCAATTATGCTGACCTACAAACCAAAACCAACCT 60

Db 1 AATTGTTATATCAAGTAGGATGGGCAATTATGCTGACCTACAAACCAAAACCAACCT 60

```
QY 61 ACCACCTTAATCAACTCCACAACCTCTTTCAGACAACCTCGTTTGTGAAAAACAATC 120
DB 61 ACCACCTTAATCAACTCCACAACCTCTTTCAGACAACCTCGTTTGTGAAAAACAATC 120
QY 121 TGTAAACAGATAACTGTGTGAAGATAACCGTGTGCGAGCCCAAAACCCGCTACTGCAACTT 180
DB 121 TGTAAACAGATAACTGTGTGAAGATAACCGTGTGCGAGCCCAAAACCCGCTACTGCAACTT 180
QY 181 TTATTGTGAACCTCCCAATTATGAAAAATCCCTTTTCGCTCTCTTCTGCTGTTTCTGTTTCTGCTT 240
DB 181 TTATTGTGAACCTCCCAATTATGAAAAATCCCTTTTCGCTCTCTTCTGCTGTTTCTGTTTCTGCTT 240
QY 241 ACTTACTGCCAGCGAAATTT 259
DB 241 ACTTACTGCCAGCGAAATTT 259

RESULT 2
US-10-311-455-1686/c
; Sequence 1686, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEFENROCK, Christian
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining Cytosine Methylation
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311.455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1686
; LENGTH: 13326
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1686

Query Match 15.9%; Score 41.2; DB 12; Length 13326;
Best Local Similarity 54.7%; Pred. No. 0.066;
Matches 82; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 36 TGACCTACAAACCAAAACCAACCTACCACCTTAATCACTCCACAAACCCCTCTTCAGA 95
DB 8480 TTATCTATAAACAATAATCAATATACCTCCCATCTCCCAACAAACCAACAGCAACTCA 8421
QY 96 CAACCTCGTTTTTGTAAAAACAATCTGTAAACAGATAACTGTGTGAAGATAACCGTTGCGG 155
DB 8420 ACATATCTCATCTCTACAAAAAACCCTCAAAATTTCTCTTTTACCAACTCCCAATTAC 8361
QY 156 AGCCCGAAACCCGCTACTGCAACTTTTATT 185
DB 8360 TACCACATAACATACCTTTAAAAATAAT 8331

RESULT 3
US-10-172-086-112/c
; Sequence 112, Application US/10172086
; Publication No. US20030113750A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Method and nucleic acids for the differentiation of prostate tumors
; TITLE OF INVENTION: of prostate tumors
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/172,086

; CURRENT FILING DATE: 2002-06-13
; NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 112
; LENGTH: 6944
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-172-086-112

Query Match 14.7%; Score 38; DB 14; Length 6944;
Best Local Similarity 54.2%; Pred. No. 0.48;
Matches 77; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 44 AAAACCAAAACAACCTTACCACCTTAATCACTCCACAACCCCTCTTTCAGACAACCTCG 103
DB 4871 AAATCCAAAAAACAACGACCTCTATAAAAAAATAAAAAACCCGCGATAAAAACGCA 4812
QY 104 TTTTGTGAAAAACAATCTGTAAACAGATAACTGTGTGAAGATAACCGTTGCGGAGCCCAA 163
DB 4811 AAACCTCAAAAAACCGCTCCGAAAAAATACTAAGAAATCTCAACCCCTCTCTCGCCCCA 4752
QY 164 AACCGTACTGCAACTTTTATT 185
DB 4751 AACTCCCACTCTTAAAAATTT 4730

RESULT 4
US-10-311-455-1293/c
; Sequence 1293, Application US/10311455
; Publication No. US20030143608A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEFENROCK, Christian
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining Cytosine Methylation
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311.455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1293
; LENGTH: 16509
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1293

Query Match 14.7%; Score 38; DB 12; Length 16509;
Best Local Similarity 53.1%; Pred. No. 0.72;
Matches 103; Conservative 0; Mismatches 90; Indels 1; Gaps 1;

QY 41 TACAAAACCAAAACAACCTTACCACCTTAATCACTCCAC-AAACCTCTTTCAGACAAC 99
DB 11520 TAAATATCAAAACTAACTACCACCTTAAATCTAAAAATGCCCAAAACCTCTTTAATACATA 11461
QY 100 CTCGTTTTTGTAAAAACAATCTGTAAACAGATAACTCTCAAGAATAACCGTTGCCGAGCC 159
DB 11460 CCTATCTAAACAAACAACAAAAATAAAAAAATAAATTTTAAATAACCAACCTTAAC 11401
QY 160 CAAAAACCCGCTACTGCAACTTTTATTGTGAACCTTCCCATTTAGAAAAATCCCTTTTCGT 219
DB 11400 TAAATCCGTTTCAAACTACTTACTAAATATTCGACCTAAAACTAATTTATCCCTAAAACT 11341
QY 220 CCTCTTTCTGTATT 233
```

```
Db 11340 TTTTITTTTAAAT 11327

RESULT 5
US-10-311-455-1232/c
; Sequence 1232, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining Cytosine Methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1232
; LENGTH: 5552
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1232

Query Match 14.6%; Score 37.8; DB 12; Length 5552;
Best Local Similarity 50.3%; Pred. No. 0.5;
Matches 93; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 48 CCAAAAACAACCTACCACCCCTTAATCACTCCACAACCCCTCTTCAGACACCTCGTTT 107
DB 4507 CAAATAAACCCCAACATCCTCTTAAAAAATCCACTTAACCTATTAAATATTACTCGTAT 4448

QY 108 TTGAAAAACAATCTGTAACAGATAACTGCTGAAGATAACGTTGCCGAGCCCAAAACC 167
DB 4447 TTTAAATCTCTTAATAATAAACAACCTCCAAACCCTTCCATTCAACCAACTTAAACCT 4388

QY 168 CGTACTGCAACTTTTATTGTGAACCTCCCATATGAGAAAATCCCTTTTGTCTCTTTTC 227
DB 4387 CAACCCCAATCAATTTCTTACACTTTTACTCTTAAAAAATCCAATCCCAATATACTTA 4328

QY 228 TGTAT 232
DB 4327 TCTTT 4323

RESULT 6
US-10-312-841-1/c
; Sequence 1, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
; FILE REFERENCE: E01/1208/WO
; CURRENT APPLICATION NUMBER: US/10/312,841
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 1
; LENGTH: 3673778
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; NAME/KEY: unsure
; LOCATION: (3294164)
US-10-312-841-1
```

```
Query Match 14.3%; Score 37; DB 12; Length 3673778;
Best Local Similarity 57.3%; Pred. No. 18;
Matches 67; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 30 TATGCGTGCCTACAAAACCAAAACCAACCTACCACCCCTTAATCACTCCAGAAACCCCTC 89
DB 3137528 TATCTTTTACCACACAAATATAAAAAAATAAACTCAAAAACTTAAAAAACTTAC 3137469

QY 90 TTCAGACAACCTCGTTTTTTTGAAAAACAATCTGTAAACAGATAACTGCTGAAGAATA 146
DB 3137468 CCAAAACCAATCACTATTATAAAACAATTATTTAAATATAACAAATTACGTATA 3137412

RESULT 7
US-10-311-455-2201/c
; Sequence 2201, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining Cytosine Methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 2201
; LENGTH: 8576
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-2201

Query Match 14.1%; Score 36.6; DB 12; Length 8576;
Best Local Similarity 51.5%; Pred. No. 1.4;
Matches 84; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 35 CTGACCTACAAACCAAAACCAACCTACCACCCCTTAATCACTCCAGAAACCCCTCTTCAG 94
DB 554 CTAATTAACAATCTCTTAACCTTTTATCAAAATCTTAACTTCTTACATTAAATTACA 495

QY 95 ACAACCTCGTTTTTTTGAAAAACAATCTGTAAACAGATAAATCTGCTGAAGAATACCGTTGCC 154
DB 494 ACATACTCCTTTAACTTAAAAAATTTAATAATCCACCTTCTTAAACCTACTTCTATC 435

QY 155 GAGCCCCAAACCCGTACTGCAACTTTTATTGTGAACITCCCCA 197
DB 434 GATTCACTCAAACTCATCTCCATCCAAATTTTATTCCTCTACTA 392

RESULT 8
US-10-311-455-2163/c
; Sequence 2163, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining Cytosine Methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
```

```

Db          588 A 588
          10
RESULT 10
US-10-311-455-1076/c
; Sequence 1076, Application US/10311455
; Publication NO. US200301436C6A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537.
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1076
; LENGTH: 7758
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1076

```

```

Query Match      13.8%; Score 35.8; DB 12; Length 7758;
Best Local Similarity 49.2%; Pred. NO. 2.4;
Matches 94; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

QY          41 TACAAAAACCAGAAACAACCTACACCCCTTAAATCAAATCCACAAAACCCCTCTTCAGACAACC 100
Db          3819 TACTAAAAAATAAATACCTTAACACAAAAATACAAAACAAAAAATAATCATCCTATATACC 3760

QY          101 TCGTTTTTTGAAAAACAATCTGTAAACAGATACTGCTGAAGAATACCGTTGCCGAGCCCC 160
Db          3759 CATCTTTCTTCTTAAACCATATTATTTTAAATTTATATATAAAAAAATAATAAAAAAAAAA 3700

QY          161 CAAAACCCCTACTGCAGACTTTTATTCTGTGAACCTCCCAATTATGAGAAAAATCCCTTTTCGTC 320
Db          3699 AAAAAAATAACAATTCACATCTCTTCTCTAAACCTCTTAAATTTTAAACAAACCATCATTTTA 3640

QY          . 221 CTCCTTCTCTGA 231
Db          3639 CCTTCTTTATA 3629

RESULT 11
US-09-205-658-43
; Sequence 43, Application US/09205658
; Patent No. US20010029617A1
; GENERAL INFORMATION:
; APPLICANT: Ruvkun, Gary
; APPLICANT: Ogg, Scott
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR
; TITLE OF INVENTION: IMPAIRED GLUCOSE TOLERANCE CONDITIONS
; FILE REFERENCE: 00786/351004
; CURRENT APPLICATION NUMBER: US/09/205,658
; CURRENT FILING DATE: 1998-12-03
; EARLIER APPLICATION NUMBER: 08/857,076
; EARLIER FILING DATE: 1997-05-15
; EARLIER APPLICATION NUMBER: 08/898,534
; EARLIER FILING DATE: 1997-07-07
; EARLIER APPLICATION NUMBER: US98/10080
; EARLIER FILING DATE: 1998-05-15
; NUMBER OF SEQ ID NOS: 328
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 43

```

```
; LENGTH: 3499
; TYPE: DNA
; ORGANISM: Caenorhabditis elegans
US-09-205-858-43

Query Match
Best Local Similarity 13.7%; Score 35.6; DB 9; Length 3499;
Matches 77; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

Qy 27 ATTATGCGTACCTACAAACCAAAACCAACCTTAACTCAACTCCACAACC 86
Db 2291 ATTTTCTTCTTCCAGATAGTAAATCATTTAGCACTGAAATTTATTGAAAAA 2350
Qy 87 CTCCTCAGACAACCTCGTTTTTGGAAAAACAATCTGTAACAGATACTGCTGAAGAATA 146
Db 2351 CTTCAATACAAATTTGTTTCGAAAAAAAATTTAAATATATATTTTCAGAAATCTT 2410
Qy 147 CCGTTGCGGAGCCCAAAACCCGTAC 172
Db 2411 CCGTCTTCATCTTTTCAAAATCCCTAC 2436

RESULT 12
US-09-844-353A-43
; Sequence 43, Application US/09844353A
; Patent No. US20020037585A1
; GENERAL INFORMATION:
; APPLICANT: Ruvkun, Gary
; APPLICANT: Kimura, Koutarou
; APPLICANT: Patterson, Garth
; APPLICANT: Ogs, Scott
; APPLICANT: Paradis, Suzanne
; APPLICANT: Tissenbaum, Heidi
; APPLICANT: Morris, Jason
; APPLICANT: Koweek, Allison
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR
; FILE OF INVENTION: IMPAIRED GLUCOSE TOLERANCE CONDITIONS
; FILE REFERENCE: 00786/351005
; CURRENT APPLICATION NUMBER: US/09/844,353A
; CURRENT FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 08/857,076
; PRIOR FILING DATE: 1997-05-15
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 43
; LENGTH: 3499
; TYPE: DNA
; ORGANISM: Caenorhabditis elegans
US-09-844-353A-43

Query Match
Best Local Similarity 13.7%; Score 35.6; DB 9; Length 3499;
Matches 77; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

Qy 27 ATTATGCGTACCTACAAACCAAAACCAACCTTAACTCAACTCCACAACC 86
Db 2291 ATTTTCTTCTTCCAGATAGTAAATCATTTAGCACTGAAATTTATTGAAAAA 2350
Qy 87 CTCCTCAGACAACCTCGTTTTTGGAAAAACAATCTGTAACAGATACTGCTGAAGAATA 146
Db 2351 CTTCAATACAAATTTGTTTCGAAAAAAAATTTAAATATATATTTTCAGAAATCTT 2410
Qy 147 CCGTTGCGGAGCCCAAAACCCGTAC 172
Db 2411 CCGTCTTCATCTTTTCAAAATCCCTAC 2436

RESULT 13
US-09-963-693-43
; Sequence 43, Application US/09963693
; Patent No. US20030181364A1
; GENERAL INFORMATION:
; APPLICANT: Ruvkun, Gary
```

```
; APPLICANT: Ogs, Scott
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR
; FILE OF INVENTION: IMPAIRED GLUCOSE TOLERANCE CONDITIONS
; FILE REFERENCE: 00786/351004
; CURRENT APPLICATION NUMBER: US/09/963,693
; CURRENT FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US/09/205,658
; PRIOR FILING DATE: 1998-12-03
; PRIOR APPLICATION NUMBER: 08/857,076
; PRIOR FILING DATE: 1997-05-15
; PRIOR APPLICATION NUMBER: 08/888,534
; PRIOR FILING DATE: 1997-07-07
; PRIOR APPLICATION NUMBER: US98/10080
; PRIOR FILING DATE: 1998-05-15
; NUMBER OF SEQ ID NOS: 328
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 43
; LENGTH: 3499
; TYPE: DNA
; ORGANISM: Caenorhabditis elegans
US-09-963-693-43

Query Match
Best Local Similarity 13.7%; Score 35.6; DB 12; Length 3499;
Matches 77; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

Qy 27 ATTATGCGTACCTACAAACCAAAACCAACCTTAACTCAACTCCACAACC 86
Db 2291 ATTTTCTTCTTCCAGATAGTAAATCATTTAGCACTGAAATTTATTGAAAAA 2350
Qy 87 CTCCTCAGACAACCTCGTTTTTGGAAAAACAATCTGTAACAGATACTGCTGAAGAATA 146
Db 2351 CTTCAATACAAATTTGTTTCGAAAAAAAATTTAAATATATATTTTCAGAAATCTT 2410
Qy 147 CCGTTGCGGAGCCCAAAACCCGTAC 172
Db 2411 CCGTCTTCATCTTTTCAAAATCCCTAC 2436

RESULT 14
US-10-240-485-46/c
; Sequence 46, Application US/10240485
; Publication No. US20030148327A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with
; FILE OF INVENTION: Metastasis
; FILE REFERENCE: 5013.1007
; CURRENT APPLICATION NUMBER: US/10/240,485
; CURRENT FILING DATE: 2002-10-02
; PRIOR APPLICATION NUMBER: PCT/EP01/03970
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 202
; SEQ ID NO 46
; LENGTH: 11996
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-485-46

Query Match
Best Local Similarity 13.7%; Score 35.6; DB 12; Length 11996;
Matches 77; Conservative 0; Mismatches 69; Indels 0; Gaps 0;
```

```

Matches 74; Conservative 0; Mismatches 64; Indels 0; Gaps 0;
Qy 40 CTACAAAACCAAAACAACTTACCACCTTAATCAACTCCACAACCCCTTTGAGACAAC 99
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1331 CTACATAAAATAAAGCGCATCATCCCTATTCTACAAAAAAATAAACAACAAAA 1272
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 100 CTCGTTTTTTGAAAAACAATCTGTAAAGAGATAACTGCTGAAGAATACCGTTGCCGAGCC 159
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1271 CCCTATATCTAAATTTAAACTTAAATACCTTAACCTCCATCTATTCTTACCACTCC 1212
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 160 CCAAAACCGTACTGCAA 177
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1211 CCACACACCCCAATCAA 1194
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 15
US-10-311-455-444/c
; Sequence 444, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311.455
; PRIOR FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 444
; LENGTH: 6534
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-444

Query Match 13.78; Score 35.4; DB 12; Length 6534;
Best Local Similarity 50.94; Pred. No. 3;
Matches 84; Conservative 0; Mismatches 81; Indels 0; Gaps 0;
Qy 32 TGCTGACCTACAAAACCAAAACCAACCTTACCACCTTAATCAACTCCACAACCCCTCTT 91
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 914 TTCGCGACCTTAAATAATCGTAACACCGAACCCCTACCTTAATCCGAACCTCTT 855
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 92 CAGACAACTCGTTTTTTGAAAAACAATCTGTAAAGAGATAACTGCTGAAGAATACCGTT 151
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 854 AAAAAAACCGCCTTTTAAATCTCTAATCTATTCAAAAAATAAACAACAAAAATAAATC 795
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 152 GCGAGCCCAAAACCGCTACTGCACTTTTATTGTGACTTCCC 196
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 794 TCCAACTATAAATCGGCTCATCTTAATATACTTATACCTCCC 750
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

Search completed: November 15, 2003, 08:32:35  
Job time : 153.767 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 14, 2003, 23:56:24 ; Search time 1053.55 Seconds  
(without alignments)  
9241.609 Million cell updates/sec

Title: US-09-928-457-96

Perfect score: 238

Sequence: 1 AATTGATACGTTGGAAAA.....ATGAAAACCTGGAGCAATT 238

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.sts.\*  
12: gb.sy.\*  
13: gb.un.\*  
14: gb.vi.\*  
15: em.ba.\*  
16: em.fun.\*  
17: em.hum.\*  
18: em.in.\*  
19: em.mu.\*  
20: em.om.\*  
21: em.or.\*  
22: em.ov.\*  
23: em.pat.\*  
24: em.ph.\*  
25: em.pl.\*  
26: em.ro.\*  
27: em.sts.\*  
28: em.un.\*  
29: em.vi.\*  
30: em.htg.hum.\*  
31: em.htg.inv.\*  
32: em.htg.other.\*  
33: em.htg.mus.\*  
34: em.htg.pln.\*  
35: em.htg.rod.\*  
36: em.htg.mam.\*  
37: em.htg.vrt.\*  
38: em.sy.\*  
39: em.htgo.hum.\*  
40: em.htgo.mus.\*  
41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	238	100.0	238	1	AF169474	AF169474 Neisseria
2	238	100.0	238	6	A68925	A68925 Sequence 96
3	238	100.0	238	6	BD063025	BD063025 DNA and s
4	238	100.0	349061	1	NM0223491	AL162753 Neisseria
5	233.2	98.0	10853	1	AE002528	AE002528 Neisseria
6	233.2	98.0	349980	6	AX044034	AX044034 Sequence
7	206.8	86.9	1155	6	AX024070	AX024070 Sequence
8	203.6	85.5	1155	6	AX024117	AX024117 Sequence
9	43.4	18.2	139189	10	AL935138	AL935138 Mouse DNA
10	43.4	18.2	195071	2	AC111049	AC111049 Mus muscu
11	43.4	18.2	318620	2	AC110910	AC110910 Mus muscu
12	43	18.1	11670	6	AX281284	AX281284 Sequence
13	43	18.1	143715	9	AL589947	AL589947 Human DNA
14	43	18.1	161828	2	AC021418	AC021418 Homo sapi
15	43	18.1	204122	2	AL450123	AL450123 Homo sapi
16	42.6	17.9	1364	8	YSCCAJ1	D21164 Saccharomyc
17	42.6	17.9	43100	8	SCB9379	U18796 Saccharomyc
18	41.8	17.6	7218	6	I66494	I66494 Sequence 14
19	41	17.2	130054	8	AC133862	AC133862 Medicago
20	41	17.2	139606	2	AC124957	AC124957 Medicago
21	41	17.2	202080	2	AC113906	AC113906 Rattus no
22	40.8	17.1	136105	9	AC023508	AC023508 Homo sapi
23	40.4	17.0	246163	2	AC121623	AC121623 Rattus no
24	40.2	16.9	128580	2	AP005466	AP005466 Oryza sat
25	40.2	16.9	158456	2	AP004731	AP004731 Oryza sat
26	40	16.8	11441	1	U67493	U67493 Methanococc
27	40	16.8	20933	6	AX458607	AX458607 Sequence
28	40	16.8	110000	6	AR271569	Continuation (11 o
29	39.8	16.7	6754	6	AX252005	AX252005 Sequence
30	39.8	16.7	6754	6	AX348778	AX348778 Sequence
31	39.8	16.7	248936	2	AC135200	AC135200 Rattus no
32	39.8	16.7	318115	2	AC098550	AC098550 Rattus no
33	39.6	16.6	12592	2	AX251796	AX251796 Sequence
34	39.6	16.6	147193	2	AC135914	AC135914 Oryza sat
35	39.6	16.6	250029	3	AE014820	AE014820 Plasmodiu
36	39.4	16.6	12901	1	AE011617	AE011617 Leptospir
37	39.4	16.6	157762	9	AP001893	AP001893 Homo sapi
38	39.4	16.6	161610	2	AC011728	AC011728 Homo sapi
39	39.4	16.6	168995	2	AP002359	AP002359 Homo sapi
40	39.4	16.6	180818	9	AP000821	AP000821 Homo sapi
41	39.2	16.5	6077	6	AX347120	AX347120 Sequence
42	39.2	16.5	110000	3	AC116984	AC116984 Dictyoste
43	39.2	16.5	330050	3	PFA929355	AL929355 Plasmodiu
44	39	16.4	145944	10	AC117259	AC117259 Mus muscu
45	39	16.4	169121	2	AC113475	AC113475 Mus muscu

# ALIGNMENTS

RESULT 1  
AF169474  
LOCUS AF169474  
DEFINITION Neisseria meningitidis strain Z2491 clone Em087 unknown sequence.  
ACCESSION AF169474  
VERSION AF169474.1 GI:9754682  
KEYWORDS  
SOURCE Neisseria meningitidis  
ORGANISM Neisseria meningitidis  
Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
Neisseriaceae; Neisseria.  
REFERENCE  
1 (bases 1 to 238)  
Perrin A., Nassif X. and Tinsley C.R.  
Identification of regions of the chromosome of Neisseria  
meningitidis and Neisseria gonorrhoeae which are specific to

238 bp DNA linear BCT 09-AUG-2000

```

pathogenic Neisseriae
Unpublished
REFERENCE 2 (bases 1 to 238)
AUTHORS Perrin,A., Nassif,X. and Tinsley,C.R.
TITLE Direct Submission
JOURNAL Submitted (16-JUL-1999) Necker Medicine Faculty, INSERM U411, 156
rue de Vaugurard, Paris 75015, France
FEATURES
    source
        Location/Qualifiers
            1..238
                /organism="Neisseria meningitidis"
                /mol_type="genomic DNA"
                /strain="22491"
                /db_xref="taxon:487"
                /clone="Em087"
BASE COUNT 85 a 18 c 64 g 71 t
ORIGIN
    Query Match 100.0%; Score 238; DB 1; Length 238;
    Best Local Similarity 100.0%; Pred. No. 2.6e-41;
    Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AATTGGATACGTTGGAAAGGGATATTTGATTGGGAATCGGATGAAGATAAGCGTAGAT 60
Db 1 AATTGGATACGTTGGAAAGGGATATTTGATTGGGAATCGGATGAAGATAAGCGTAGAT 60
Qy 61 GAGTTGGGAAAAAAGCTTTAGACATATCGGTAAAGATGAACCGTTATTTGTTGAAAAAT 120
Db 61 GAGTTGGGAAAAAAGCTTTAGACATATCGGTAAAGATGAACCGTTATTTGTTGAAAAAT 120
Qy 121 CTACTGGTTAACTTCAATCAGGAAAAACATCAAGAAAGTTAGGAAGTTGATTATCAGTTG 180
Db 121 CTACTGGTTAACTTCAATCAGGAAAAACATCAAGAAAGTTAGGAAGTTGATTATCAGTTG 180
Qy 181 ATAGAGTTAGATTTTCTGGAACCTTTTGTGAGGATTTCTATGAAAACTGGAAGCAATT 238
Db 181 ATAGAGTTAGATTTTCTGGAACCTTTTGTGAGGATTTCTATGAAAACTGGAAGCAATT 238
RESULT 2
LOCUS A68925 238 bp DNA linear PAT 06-MAY-1999
DEFINITION Sequence 96 from Patent W09802547.
ACCESSION A68925
VERSION A68925.1 GI:4759844
KEYWORDS
SOURCE
ORGANISM
    Neisseria meningitidis
    unclassified.
REFERENCE 1 (bases 1 to 238)
AUTHORS Nassif,X., Tinsley,C., Achtman,M., Ruelle,J.L., Vinals,C. and
Merker,P.
TITLE DNA AND SPECIFIC PROTEINS OR PEPTIDES OF THE NEISSERIA MENINGITIDIS
SPECIES BACTERIA, METHOD FOR OBTAINING THEM AND THEIR BIOLOGICAL
APPLICATIONS
JOURNAL Patent: JP 2001504684-A 87 10-APR-2001;
INSTITUT NATIONAL DE LA SANTE ET DE LA RECHERCHE MEDICALE, MAX
PLANCK GESELLSCHAFT ZUR FORDERUNG DER WISSENSCHAFTEN EV BERLIN,
SMITHKLINE BEECHAM
PN JP 2001504684-A/87
PD 10-APR-2001
PF 11-JUL-1997 JP 19980505685
PR 12-JUL-1996 FR 96/08768
PI XAVIER NASSIF, COLIN TINSLEY, MARK ACHTMAN, JEAN LOUIS RUELLE, PI
CARLA VINALS,
PI PETRA MERKER
PC C12N15/31, C07K14/22, C07K16/12, A61K39/095, C12Q1/68, G01N33/53 CC
Strandedness: Single;
CC Topology: Linear;
FH Key Location/Qualifiers.
FEATURES
    source
        Location/Qualifiers
            1..238
                /organism="unidentified"
                /mol_type="genomic DNA"
                /db_xref="taxon:32644"
BASE COUNT 85 a 18 c 64 g 71 t
ORIGIN
    Query Match 100.0%; Score 238; DB 6; Length 238;
    Best Local Similarity 100.0%; Pred. No. 2.6e-41;
    Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AATTGGATACGTTGGAAAGGGATATTTGATTGGGAATCGGATGAAGATAAGCGTAGAT 60
Db 1 AATTGGATACGTTGGAAAGGGATATTTGATTGGGAATCGGATGAAGATAAGCGTAGAT 60
Qy 61 GAGTTGGGAAAAAAGCTTTAGACATATCGGTAAAGATGAACCGTTATTTGTTGAAAAAT 120
Db 61 GAGTTGGGAAAAAAGCTTTAGACATATCGGTAAAGATGAACCGTTATTTGTTGAAAAAT 120
Qy 121 CTACTGGTTAACTTCAATCAGGAAAAACATCAAGAAAGTTAGGAAGTTGATTATCAGTTG 180
Db 121 CTACTGGTTAACTTCAATCAGGAAAAACATCAAGAAAGTTAGGAAGTTGATTATCAGTTG 180
Qy 181 ATAGAGTTAGATTTTCTGGAACCTTTTGTGAGGATTTCTATGAAAACTGGAAGCAATT 238
Db 181 ATAGAGTTAGATTTTCTGGAACCTTTTGTGAGGATTTCTATGAAAACTGGAAGCAATT 238

```

```

pathogenic Neisseriae
Unpublished
REFERENCE 2 (bases 1 to 238)
AUTHORS Perrin,A., Nassif,X. and Tinsley,C.R.
TITLE Direct Submission
JOURNAL Submitted (16-JUL-1999) Necker Medicine Faculty, INSERM U411, 156
rue de Vaugurard, Paris 75015, France
FEATURES
    source
        Location/Qualifiers
            1..238
                /organism="Neisseria meningitidis"
                /mol_type="genomic DNA"
                /strain="22491"
                /db_xref="taxon:487"
                /clone="Em087"
BASE COUNT 85 a 18 c 64 g 71 t
ORIGIN
    Query Match 100.0%; Score 238; DB 1; Length 238;
    Best Local Similarity 100.0%; Pred. No. 2.6e-41;
    Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AATTGGATACGTTGGAAAGGGATATTTGATTGGGAATCGGATGAAGATAAGCGTAGAT 60
Db 1 AATTGGATACGTTGGAAAGGGATATTTGATTGGGAATCGGATGAAGATAAGCGTAGAT 60
Qy 61 GAGTTGGGAAAAAAGCTTTAGACATATCGGTAAAGATGAACCGTTATTTGTTGAAAAAT 120
Db 61 GAGTTGGGAAAAAAGCTTTAGACATATCGGTAAAGATGAACCGTTATTTGTTGAAAAAT 120
Qy 121 CTACTGGTTAACTTCAATCAGGAAAAACATCAAGAAAGTTAGGAAGTTGATTATCAGTTG 180
Db 121 CTACTGGTTAACTTCAATCAGGAAAAACATCAAGAAAGTTAGGAAGTTGATTATCAGTTG 180
Qy 181 ATAGAGTTAGATTTTCTGGAACCTTTTGTGAGGATTTCTATGAAAACTGGAAGCAATT 238
Db 181 ATAGAGTTAGATTTTCTGGAACCTTTTGTGAGGATTTCTATGAAAACTGGAAGCAATT 238
RESULT 3
LOCUS BD063025 238 bp DNA linear PAT 27-AUG-2002
DEFINITION DNA and specific proteins or peptides of the Neisseria meningitidis
species bacteria, method for obtaining them and their biological
applications.
ACCESSION BD063025
VERSION BD063025.1 GI:22608628
KEYWORDS JP 2001504684-A/87.
SOURCE unidentified
ORGANISM unidentified
unclassified.
REFERENCE 1 (bases 1 to 238)
AUTHORS Nassif,X., Tinsley,C., Achtman,M., Ruelle,J.L., Vinals,C. and
Merker,P.
TITLE DNA and specific proteins or peptides of the Neisseria meningitidis
species bacteria, method for obtaining them and their biological
applications
JOURNAL Patent: JP 2001504684-A 87 10-APR-2001;
INSTITUT NATIONAL DE LA SANTE ET DE LA RECHERCHE MEDICALE, MAX
PLANCK GESELLSCHAFT ZUR FORDERUNG DER WISSENSCHAFTEN EV BERLIN,
SMITHKLINE BEECHAM
PN JP 2001504684-A/87
PD 10-APR-2001
PF 11-JUL-1997 JP 19980505685
PR 12-JUL-1996 FR 96/08768
PI XAVIER NASSIF, COLIN TINSLEY, MARK ACHTMAN, JEAN LOUIS RUELLE, PI
CARLA VINALS,
PI PETRA MERKER
PC C12N15/31, C07K14/22, C07K16/12, A61K39/095, C12Q1/68, G01N33/53 CC
Strandedness: Single;
CC Topology: Linear;
FH Key Location/Qualifiers.
FEATURES
    source
        Location/Qualifiers
            1..238
                /organism="unidentified"
                /mol_type="genomic DNA"
                /db_xref="taxon:32644"
BASE COUNT 85 a 18 c 64 g 71 t
ORIGIN
    Query Match 100.0%; Score 238; DB 6; Length 238;
    Best Local Similarity 100.0%; Pred. No. 2.6e-41;
    Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AATTGGATACGTTGGAAAGGGATATTTGATTGGGAATCGGATGAAGATAAGCGTAGAT 60
Db 1 AATTGGATACGTTGGAAAGGGATATTTGATTGGGAATCGGATGAAGATAAGCGTAGAT 60
Qy 61 GAGTTGGGAAAAAAGCTTTAGACATATCGGTAAAGATGAACCGTTATTTGTTGAAAAAT 120
Db 61 GAGTTGGGAAAAAAGCTTTAGACATATCGGTAAAGATGAACCGTTATTTGTTGAAAAAT 120
Qy 121 CTACTGGTTAACTTCAATCAGGAAAAACATCAAGAAAGTTAGGAAGTTGATTATCAGTTG 180
Db 121 CTACTGGTTAACTTCAATCAGGAAAAACATCAAGAAAGTTAGGAAGTTGATTATCAGTTG 180
Qy 181 ATAGAGTTAGATTTTCTGGAACCTTTTGTGAGGATTTCTATGAAAACTGGAAGCAATT 238
Db 181 ATAGAGTTAGATTTTCTGGAACCTTTTGTGAGGATTTCTATGAAAACTGGAAGCAATT 238

```



```

RESULT 4
NMA222491
LOCUS
DEFINITION Neisseria meningitidis serogroup A strain Z2491 complete genome;
segment 2/7.
ACCESSION AL162753 AL157959
VERSION AL162753.2 GI:7379120
KEYWORDS
SOURCE
ORGANISM
Neisseria meningitidis Z2491
Neisseria meningitidis Z2491
Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
Neisseriaceae; Neisseria.
1 (bases 1 to 349061)
Parkhill,J., Achtman,M., James,K.D., Bentley,S.D., Churcher,C.,
Klee,S.R., Morelli,G., Basham,D., Brown,D., Chillingworth,T.,
Davies,R.M., Davis,P., Devlin,K., Feltwell,T., Hamlin,N.,
Holroyd,S., Jagels,K., Leather,S., Moule,S., Mungall,K.,
Quail,M.A., Rajandream,M.A., Rutherford,K.M., Simmonds,M.,
Skellton,J., Whitehead,S., Spratt,B.G. and Barrall,B.G.
Complete DNA sequence of a serogroup A strain of Neisseria
meningitidis Z2491
Nature 404 (6777), 502-506 (2000)
JOURNAL
MEDLINE 20222556
PUBMED 10761919
REFERENCE 2 (bases 1 to 349061)
AUTHORS Parkhill,J.
TITLE Direct Submission
SUBMITTED (30-MAR-2000) Submitted on behalf of the Neisseria
sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk
NOTES:
Details of N. meningitidis sequencing at the Sanger Centre are
available on the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/N_meningitidis/).
FEATURES
source
1. 349061
/organism="Neisseria meningitidis Z2491"
/mol_type="genomic DNA"
/strain="Z2491"
/db_xref="taxon:122587"
/notes="serogroup: A"
complement(24..206)
repeat_unit
/notes="ATR repeat; hmms hit to HMM ATR (1 - 183), score:
310.39"
label=ATR
209..212
/genes="NMA0368"
220..681
/genes="NMA0368"
220..681
/genes="NMA0368"
/notes="NMA0368, probable integral membrane protein, len:
153 aa; contains four probable transmembrane domains"
/codon_start=1
/transl_table=11
/product="putative integral membrane protein"
/protein_id="CAB83669.1"
/db_xref="GI:7379121"
/db_xref="SPTREMBL:Q9JWH7"
/translation="MQEQNRKSPPIVMLLVSVLWLTASLSNVAFYLGNGHSGEGLTV
LILGSI PASIDIRYCAVYANYVWLAAIVLLALRKVPVPHAAFWGLALVAFSVKAVYV
DEAGNTSDIVRYGAGFYLVYAAFAVASIGTFAGNKKERKAASADGTKNDV"
671..1492
/genes="hemK"
671..1492
/genes="hemK"
/notes="NMA0369, hemK, HemK protein, len: 273 aa; similar
to e.g. HEMK_ECOLI_P37186 HEMK protein (277 aa), fasta
scores; E(): 0, 42.3% identity in 279 aa overlap. Contains
PS00092 N-6 Adenine-specific DNA methylases signature"
/codon_start=1
/transl_table=11

```

```

/product="HemK protein"
/protein_id="CAB83670.1"
/db_xref="GI:7379122"
/db_xref="SPTREMBL:Q9JWH5"
/translation="MTEDKWLGLSKLPKNEARMLLOYSEYTRVOLLTRGGEMPEDEV
RGRARLQRRLNGEPVAYILIGRFFYGRFVNFSLVLPREPETHLVEAVLARLPEN
GRVLDLGTSGAVAVVALERPDADFRASDISPPALETARKVAADLGARVEFYGSWF
DTDMPSEKWDIIVSNPYIENGDLKSQLRFPETALDFFDGLSLCIRTLAQAP
DLRARGFLLEHGFQDGAARVGLAENGFSVETLPDLAGLDVRTLKGKMKLK"
831..840
/misc_feature
/genes="hemK"
/notes="Core DNA uptake sequence: gcgcgtctgaa"
/label=DUS
1162..1171
/genes="hemK"
/notes="Core DNA uptake sequence: gcgcgtctgaa"
/label=DUS
1187..1207
/misc_feature
/genes="hemK"
/notes="PS00092 N-6 Adenine-specific DNA methylases
signature"
complement(1279..1288)
/notes="Core DNA uptake sequence: gcgcgtctgaa"
/label=DUS
1565..1568
RBS
1578..1596
gene
/genes="NMA0370"
1578..1596
CDS
/genes="NMA0370"
/notes="NMA0370, probable integral membrane protein, len:
462 aa; similar to hypothetical proteins e.g. Y325_HAEIN
P44640 hypothetical protein HI0325 (450 aa), fasta scores;
E(): 0, 49.8% identity in 464 aa overlap"
/codon_start=1
/transl_table=11
/product="putative integral membrane protein"
/protein_id="CAB83671.1"
/db_xref="GI:7379123"
/db_xref="SPTREMBL:Q9JWH5"
/translation="NNAVVAVIVMLVLSLRVHVVLSTLTGAFVGGAVAGMPLQNI
A
DAAGVQAGIIPFVKGLLEGAKIALSYMIGAFAMATHTSGLPQQAGAVVRKLNR
GMGPDVSRGEGVWKLLSLILVNMNSQNIPIHAFIPMIVPLLVFNRKLDR
RLIACVITPGLVTFYFPGFALFELNELLGNHSAAPQDLDKVINVAAMAIPAL
GMLAGLLAFVHYRPRLYQSNNAADAGNADANRPPQSAVRSAAVAIAVCFAIQL
MYEDSLVLGMLGEAVFMWLVNEDKADVFEGEIKMWAVGFTMIAAQGFAAVMNA
TCHIQPLVSSMAIFGNSKGAALAMLVGLLVMTGIGSSSESTLIIAIIYVPLCVGL
GFSPLATVAIVGTAGALDAGSPASDSTLGTPTMGNADQCHDHIRDSVLPITFIHNP
LLIAGWIAAMVL"
2967..3191
gene
/genes="slyX"
2967..3191
CDS
/genes="slyX"
/notes="NMA0371, slyX, SLIX protein homolog, len: 74 aa;
similar to SLIX_HAEIN_P44759 SLIX protein homolog (73 aa),
fasta scores; E(): 0.0017, 33.8% identity in 74 aa
overlap, and SLIX_ECOLI_P30857 SLIX protein (72 aa), fasta
scores; E(): 0.24, 32.4% identity in 68 aa overlap"
/codon_start=1
/transl_table=11
/product="SLIX protein homolog"
/protein_id="CAB83672.1"
/db_xref="GI:7379124"
/db_xref="SPTREMBL:Q9JWH4"
/translation="MDAVQEPHRTLEIQSALQEDVIAGLNAMVAELRQTLDLQQA
QLRLLYQKMQDENPDQAQBPVSLRDEIPHY"
Complement(3271..3423)
/genes="NMA0372"
complement(3271..3423)
/genes="NMA0372"
/notes="NMA0372, unknown, questionable CDS, len: 50 aa"
/codon_start=1
/transl_table=11
/product="very hypothetical protein NMA0372"

```

```

Db      326855 CTACTGGTTAACTTCATCAGGAAACATGAGAAAGTAGGAAAGTTGATTATTCAGTTG 326914
QY      181 ATAGAGTTAGATTCTTCTGGAACCTTTGTGAGGAGATTCTATGAAAACTGGAAGCAATT 238
       |||||
Db      326915 ATAGAGTTAGATTCTTCTGGAACCTTTGTGAGGAGATTCTATGAAAACTGGAAGCAATT 326972
       |||||

RESULT 5
AE002528      10853 bp      DNA      linear      BCT 25-MAY-2000
LOCUS      Neisseria meningitidis serogroup B strain MC58 section 170 of 206
DEFINITION      of the complete genome.
ACCESSION      AE002528 AE002038
VERSION      AE002528.1 GI:7227034
KEYWORDS
SOURCE      Neisseria meningitidis MC58
ORGANISM      Neisseria meningitidis MC58
REFERENCE      1 (bases 1 to 10853)
AUTHORS      Tettelin, H., Saunders, N.J., Heidelberg, J., Jeffries, A.C.,
              Nelson, K.E., Eisen, J.A., Ketchum, K.A., Hood, D.W., Peden, J.F.,
              Dodson, R.J., Nelson, W.C., Gwinn, M.L., DeBoy, R., Peterson, J.D.,
              Hickey, E.K., Haft, D.H., Salzberg, S.L., White, O., Fleischmann, R.D.,
              Dougherty, B.A., Mason, T., Ciecko, A., Parksey, D.S., Blait, E.,
              Citron, H., Clark, E.B., Cotton, M.D., Uterback, T.R., Khouri, H.,
              Qian, H., Vamathevan, J., Gill, J., Scarlato, V., Masignani, V.,
              Pizzu, M., Grandi, G., Sun, L., Smith, H.O., Fraser, C.M., Moxon, E.R.,
              Rappuoli, R. and Venter, J.C.
              Complete genome sequence of Neisseria meningitidis serogroup B
              strain MC58
JOURNAL      Science 287 (5459), 1809-1815 (2000)
MEDLINE      20175755
PUBMED      10710307
REFERENCE      2 (bases 1 to 10853)
AUTHORS      Tettelin, H., Saunders, N.J., Heidelberg, J., Jeffries, A.C.,
              Nelson, K.E., Eisen, J.A., Ketchum, K.A., Hood, D.W., Peden, J.F.,
              Dodson, R.J., Nelson, W.C., Gwinn, M.L., DeBoy, R., Peterson, J.D.,
              Hickey, E.K., Haft, D.H., Salzberg, S.L., White, O., Fleischmann, R.D.,
              Dougherty, B.A., Mason, T., Ciecko, A., Parksey, D.S., Blait, E.,
              Citron, H., Clark, E.B., Cotton, M.D., Uterback, T.R., Khouri, H.,
              Qian, H., Vamathevan, J., Gill, J., Scarlato, V., Masignani, V.,
              Pizzu, M., Grandi, G., Sun, L., Smith, H.O., Fraser, C.M., Moxon, E.R.,
              Rappuoli, R. and Venter, J.C.
              Direct Submission
TITLE      Submitted (17-MAR-2000) The Institute for Genomic Research, 9712
JOURNAL      Medical Center Dr, Rockville, MD 20850, USA
FEATURES      Location/Qualifiers
              source
              1..10853
              /organism="Neisseria meningitidis MC58"
              /mol_type="genomic DNA"
              /strain="MC58"
              /db_xref="taxon:122586"
              /note="serogroup: B"
              /codon_start=1
              /complement(85..1827)
              /gene="NM01780"
              /protein_id="AAF42120.1"
              /note="similar to PID:1621558 percent identity: 98.55;
              identified by sequence similarity; putative"
              /trans_table=11
              /product="Hemolysin activation protein HecB, putative"
              /protein_id="AAF42120.1"
              /db_xref="GI:7227035"
              /translation="MKFFPAPCLIVLAVILPKLTAAENDABELIRSMQQQHIDABL
              MLDGNLRLQKAAQOILVRGLYTSQALIQPNMDSGLIKLRVAGSGIDIRYEEK
              LIDANVFPEQLEKNVLSDEPTCRVNYISLDKTVKFSFLPSVLMKETAFKFG
              LTCGSAAGSISAFNNKFLYRNKLINRDVEQGLNLRPLPSYKTDQIIPSEEGS
              DLQIKQONKPIRFSIGDIDGAGTKYKGNVALSFDNPLGDLDFVYIGRGLAHK
              TDLDTKGTETESGSRYSVHYSPVKWLFSEFNHNRVHEATGEGSVVNDYGRKQY
              QSSLAARMLRNRLHRTSVGMKLWTKYKIDDAIEIVORRSAGWEALRHRYAL
              NRWLDGKLSYKRGTMGRQSMPEENGDDILPQTRMKITITASLDAAPFILGKQOF

misc_feature
       /protein_id="CAB83673.1"
       /db_xref="GI:7379125"
       /db_xref="SPTREMBL:Q9JWH3"
       /translation="MFGRSLPLGSDGFLLAVVEYVPVFGICLQKTA YFRSGGRKFRHRI
       KGILSD"
       3562..3571
       /note="Core DNA uptake sequence: gccgtctgaa"
       /label=DUS
       complement(3586..4356)
       /gene="thiF"
       /gene="thiF"
       /note="NMA0373, thiF, probable ThiF protein, len: 256 aa;
       similar to e.g. THIF_ECOLI P30138 ThiF protein (251 aa),
       fasta scores: E(): 0, 43.1% identity in 246 aa overlap,
       and MOBB_ECOLI P12282 molybdoxin biosynthesis MOBB
       protein (249 aa), fasta scores: E(): 0, 43.9% identity in
       244 aa overlap (note that N.m. does not have orthologs of
       any other molybdoxin biosynthesis proteins). Contains
       pfam match to entry PF00899 ThiF_family, ThiF family"
       /codon_start=1
       /trans_table=11
       /product="ThiF protein"
       /protein_id="CAB83674.1"
       /db_xref="GI:7379126"
       /db_xref="SPTREMBL:Q9JWH2"
       /translation="MTTENDDDAFLRYSRHILLDEIGECOKLSAAHILVVGCG
       GLGAALPYAAGSGITLADSVLDELHNLQKQVAFDEGVGKLTALADLRHIN
       HTVDVTRINKLDCRGTIGVQAADIVLDCDNYATRVACVQAKTLPVSGAAVR
       FEGCLAYRPLDPSQCYALCDGGSAGCSLFGVFSPLGIIIGSTQAAELKILL
       DAGBPSHGLAVRALEGQVQYFDLPNPECPVCGAER"
       3787..3796
       /note="Core DNA uptake sequence: gccgtctgaa"
       /label=DUS
       complement(3847..4257)
       /gene="thiF"
       /note="Upfam match to entry PF00899 ThiF_family, ThiF
       family, score 186.60, E-value 4.1e-52"
       complement(4418..4427)
       /note="Core DNA uptake sequence: gccgtctgaa"
       /label=DUS
       4471..7224
       /gene="ppc"
       4471..7224
       /gene="ppc"
       /gene="ppc"
       /EC_number="4.1.1.31"
       /note="NMA0374, ppc, phosphoenolpyruvate carboxylase, len:
       917 aa; similar to many e.g. CAPP_RHOPA O3483
       phosphoenolpyruvate carboxylase (EC 4.1.1.31) (936 aa),
       fasta scores: E(): 0, 43.3% identity in 928 aa overlap.
       Contains 2x Pfam match to entry PF00311 PEPcase,
       Phosphoenolpyruvate carboxylase, PS00017 ATP/GTP-binding
       site motif A (P-loop), and PS00393 Phosphoenolpyruvate
       carboxylase active site 2"
       /codon_start=1
       /trans_table=11
       /product="phosphoenolpyruvate carboxylase"
       /protein_id="CAB83675.1"
       100.0%; Score 238; DB 1; Length 349061;
       Best Local Similarity 100.0%; Pred. No. 6,7e-42;
       Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AATTGGATACGTTGGAAAAGGATATTTGATTGGGAATGGGATGAAGTAAACGCTAGAT 60
Db      326735 AATTGGATACGTTGGAAAAGGATATTTGATTGGGAATGGGATGAAGTAAACGCTAGAT 326794
QY      61 GACTTGGGAAAAGCTTAGACATATCGGTAAAGCAACCGCTATTGTGAAAAT 120
Db      326795 GAGTTGGGAAAAGCTTAGACATATCGGTAAAGCAACCGCTATTGTGAAAAT 326854
QY      121 CTACTGTTAACTTCATCAGGAAACATGAGAAAGTAGGAAAGTTGATTATTCAGTTG 180

```



```

RESULT 6
AX044034/c
LOCUS       AX044034       349980 bp       DNA       linear       PAT 24-NOV-2000
DEFINITION   Sequence 113 from Patent WO066791.
ACCESSION   AX044034
VERSION     AX044034.1   GI:11342918
KEYWORDS
SOURCE      Neisseria meningitidis
ORGANISM    Neisseria meningitidis
            Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
            Neisseriaceae; Neisseria.
REFERENCE   1
AUTHORS     Pizza, M., Hickey, E., Peterson, J., Tettelin, H., Venter, J.C.,
            Masignani, V., Galeotti, C., Mora, M., Ratti, G., Scarselli, M.,
            Scarlato, V., Rappelli, R., Frazer, C.M. and Grandi, G.
TITLE       Neisseria genomic sequences and methods of their use
JOURNAL     Patent: WO 0066791-A 113 09-NOV-2000;
            CHIRON CORPORATION (US); THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES
            Location/Qualifiers
            source          1..349980
            /organism="Neisseria meningitidis"
            /mol_type="genomic DNA"
            /db_xref="taxon:487"
            /note="sequence too long, cut in 8 pieces -seq 1: 1 to
            349980 bases-seq 108: 300001 to 649980 349980
            bases-seq 109: 600001 to 949980 349980 bases-seq 110:
            900001 to 1249980 349980 bases-seq 111: 1200001 to 1549980
            349980 bases-seq 112: 1500001 to 1849980 349980 bases-seq
            113: 1800001 to 2149980 349980 bases-seq 114: 2100001 to
            2272325 172325 bases"
BASE COUNT   86473 a 35646 c 85908 g 81953 t
ORIGIN
Query Match      98.0%; Score 233.2; DB 6; Length 349980;
Best Local Similarity 98.7%; Pred. No. 7.2e-41;
Matches 235; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AATTGGATACGTTGGAAAGGGATATTGATTCGGGATCGGATGAGTAACCGTAGAT 60
|||
Db 72678 AATTGGATACGTTGGAAAGGGATATTGATTCGGGATCGGATGAGTAACCGTAGAT 72619

QY 61 GAGTTGGGAAAAAAGTTAGAACATATCGGTAAGAAATCAACCGTTATTGTTGAAAAAT 120
|||
Db 72618 GAGTTGGGAAAAAAGTTAGAACATATCGGTAAGAAATCAACCGTTATTGTTGAAAAAT 72559

QY 121 CTACTGGTTAACTCAATCAGGAAAAACATGAAAGTTAGGAAGTTGATTCAGTTG 180
|||
Db 72558 CTACTGGTTAACTCAATCAGGAAAAATATGAAAGTTAGGAAGTTGATTCAGTTG 72499

QY 181 ATAGAGTTAGATTTCTGGAACCTTTCTGAGGGATTCATGAAAACTGGAAGCAATT 238
|||
Db 72498 ATAGAGTTAGATTTCTGGAACCTTTCTGAGGGATTCATGAAAACTGGAAGCAATT 72441

RESULT 7
AX024070
LOCUS       AX024070       1155 bp       DNA       linear       PAT 15-SEP-2000
DEFINITION   Sequence 13 from Patent FR2785293.
ACCESSION   AX024070
VERSION     AX024070.1   GI:10184382
KEYWORDS
SOURCE      Neisseria meningitidis
ORGANISM    Neisseria meningitidis
            Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
            Neisseriaceae; Neisseria.
REFERENCE   1
AUTHORS     Nassif, X., Tinsley, C., Aujame, L., Perrin, A., Rokbi, B.,
            Bouchardon, A. and Renaud, M.G.
TITLE       Neisseria meningitidis
JOURNAL     Patent: FR 2785293-A 13 05-MAY-2000;
            PASTEUR MERIEUX SERUMS VACC (FR)
FEATURES
            Location/Qualifiers
            source          1..1155
            /organism="Neisseria gonorrhoeae"
            /mol_type="genomic DNA"
            /db_xref="taxon:485"
            /note="unnamed protein product"
            /codon_start=1
            /transl_table=11
            /protein_id="CAC08934.1"
            /db_xref="GI:10184429"
            /translation="WGHLDPGIGPKTFROTLYQKPKFKGAVRNLEAASCKYINEL
            YQRADPTAPLFLKRGKGAIVPKKEIVSFDDLGKTRVRFKSVIYEHKNGASLVYNH
            INNEPFDHIAQVARFAGAHITVSGYLAQSDSYKHNWDTRDVAIQLFGRKRWQL
            TAPDFPMLYMQQTKDIDISPEIDMDIILEAGDVLYIPRGWHRHRIPIGCEPFFHA
            VGTFFPNGYNLYEWMKFKFTIESLRHSFSDWEQDRTINDTAAQIAAMIAADPVNYEA
            DELGKKVLEHIGKNEPLLKLLNLFNFGKHEEVKRLIYQIIELDLFLEL"
BASE COUNT   335 a 269 c 248 g 303 t
ORIGIN
Query Match      86.9%; Score 206.8; DB 6; Length 1155;
Best Local Similarity 99.0%; Pred. No. 9.8e-35;
Matches 208; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AATTGGATACGTTGCAAAAGGGATATTGATTCGGGATCGGATGAGTAACCGTAGAT 60
|||
Db 946 AATTGGATACGTTGCAAAAGGGATATTGATTCGGGATCGGATGAGTAACCGTAGAT 1005

QY 61 GAGTTGGGAAAAAAGTTAGAACATATCGGTAAGAAATCAACCGTTATTGTTGAAAAAT 120
|||
Db 1006 GAAATTGGGAAAAAAGTTAGAACATATCGGTAAGAAATCAACCGTTATTGTTGAAAAAT 1065

QY 121 CTACTGGTTAACTCAATCAGGAAAAACATGAAAGTTAGGAAGTTGATTCAGTTG 180
|||
Db 1056 CTACTGGTTAACTCAATCAGGAAAAACATGAAAGTTAGGAAGTTGATTCAGTTG 1125

QY 181 ATAGAGTTAGATTTCTGGAACCTTTCTGGAACCTTTGTA 210
|||
Db 1126 ATAGAGTTAGATTTCTGGAACCTTTCTGGAACCTTTGTA 1155

RESULT 8
AX024117
LOCUS       AX024117       1155 bp       DNA       linear       PAT 15-SEP-2000
DEFINITION   Sequence 60 from Patent FR2785293.
ACCESSION   AX024117
VERSION     AX024117.1   GI:10184428
KEYWORDS
SOURCE      Neisseria gonorrhoeae
ORGANISM    Neisseria gonorrhoeae
            Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
            Neisseriaceae; Neisseria.
REFERENCE   1
AUTHORS     Nassif, X., Tinsley, C., Aujame, L., Perrin, A., Rokbi, B.,
            Bouchardon, A. and Renaud, M.G.
TITLE       Neisseria gonorrhoeae
JOURNAL     Patent: FR 2785293-A 60 05-MAY-2000;
            PASTEUR MERIEUX SERUMS VACC (FR)
FEATURES
            Location/Qualifiers
            source          1..1155
            /organism="Neisseria gonorrhoeae"
            /mol_type="genomic DNA"
            /db_xref="taxon:485"
            /note="unnamed protein product"
            /codon_start=1
            /transl_table=11
            /protein_id="CAC08934.1"
            /db_xref="GI:10184429"
            /translation="WGHLDPGIGPKTFROTLYQKPKFKGAVRNLEAASCKYINEL
            YQRADPTAPLFLKRGKGAIVPKKEIVSFDDLGKTRVRFKSVIYEHKNGASLVYNH
            INNEPFDHIAQVARFAGAHITVSGYLAQSDSYKHNWDTRDVAIQLFGRKRWQL
            TAPDFPMLYMQQTKDIDISPEIDMDIILEAGDVLYIPRGWHRHRIPIGCEPFFHA
            VGTFFPNGYNLYEWMKFKFTIESLRHSFSDWEQDRTINDTAAQIAAMIAADPVNYEA
            DELGKKVLEHIGKNEPLLKLLNLFNFGKHEEVKRLIYQIIELDLFLEL"
BASE COUNT   335 a 269 c 248 g 303 t
ORIGIN

```

FSDFPLGKERTDTAFHLEQFANPNATPLSDVRLRLNANNLDLTKYLIQNGMKISV  
DELGGKVLHIGKREPULLKNLVNFQAKHEEVKRLIYOLDFLEIL"

BASE COUNT  
ORIGIN

337 a 265 c 246 g 307 t

Query Match 85.5%; Score 203.6; DB 6; Length 1155;  
Best Local Similarity 98.1%; Pred. No. 4.8e-34; Indels 0; Gaps 0;  
Matches 206; Conservative 0; Mismatches 4;  
QY 1 AATTGGATAGCTGGAAAAGGGATATTGTTGGGAATGGATCAAGATAAGCGTAGAT 60  
DB AATTGGATAGCTGGAAAAGGGATATTGTTGGGAATGGATCAAGATAAGCGTAGAT 1005  
QY 61 GAGTTGGGAAAAAGTTTGAACATATCGTGAAGATGAACCGTTATTGTTGAAAAAT 120  
DB GAGTTGGGAAAAAGTTTGAACATATCGTGAAGATGAACCGTTATTGTTGAAAAAT 1065  
QY 121 CTACTGGTTAACTTCAATCAGGAAAAACATCAAGAAGTTAGCAAGTTGATTATCAGTTG 180  
DB CTACTGGTTAACTTCAATCAGGAAAAACATCAAGAAGTTAGCAAGTTGATTATCAGTTG 1125  
QY 181 ATAGAGTTAGATTTCTCGAACTTTGTGA 210  
DB ATAGAGTTAGATTTCTCGAAATTTGTGA 1155

RESULT 9  
AL935138

LOCUS Mouse DNA sequence from clone RP23-222H15 on chromosome 2, complete  
DEFINITION

AL935138 139189 bp DNA linear ROD 24-DEC-2002

ACCESSION AL935138

VERSION AL935138.7 GI:27374701

KEYWORDS HTG.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (bases 1 to 139189)

Clark, S.

Direct Submission

Submitted (21-DEC-2002) Wellcome Trust Sanger Institute, Hinxton,

Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

On Dec 26, 2002 this sequence version replaced GI:26985395.

Sequence from the Mouse Genome Sequencing Consortium whole genome

shotgun may have been used to confirm this sequence. Sequence data

from the whole genome shotgun alone has only been used where it has

a Phred quality of at least 30.

----- Genome Center

Center: Wellcome Trust Sanger Institute

Web site: <http://www.sanger.ac.uk>

Contact: humquery@sanger.ac.uk

-----

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at

<http://www.sanger.ac.uk/Projects/C.elegans/wormpep> RP23-222H15 is from the RP23-23 Mouse PAC Library constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm> VECTOR: pBACE3.6

FEATURES  
source

Location/Qualifiers  
1..139189  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="2"  
/clone="RP23-222H15"  
/clone\_lib="RP23-23"

BASE COUNT 42836 a 24714 c 25557 g 46082 t  
ORIGIN

Query Match 18.2%; Score 43.4; DB 10; Length 139189;  
Best Local Similarity 64.4%; Pred. No. 4.8;  
Matches 65; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 118 AATCTACTGGTTAACTTCAATCAGGAAAAACATCAAGAAGTTAGCAAGTTGATTATCAG 177  
DB 89602 AATCTATTTTAAATTTTATATCTTAAAGCCTTAAGCAATGCAAAATGTTTATAAG 89661  
QY 178 TTGATAGAGTTAGATTTCTGGAACCTTTGTGAGGATTTCT 218  
DB 89662 TTATATAGATAGAAATTAAGAGCTTTTGTAAAAATTTCT 89702

RESULT 10

AC111049

LOCUS

DEFINITION

AC111049 195071 bp DNA linear HTG 31-OCT-2002

pieces.

AC111049

AC111049.3 GI:24431747

KEYWORDS HTG; HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 195071)

Birren, B., Nusbaum, C. and Lander, E.

Mus musculus, clone RP23-421C14

Unpublished

2 (bases 1 to 195071)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,

Anderson, S., Barna, N., Bastien, V., Boguski, L., Boukhgalter, B.,

Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,

Choquel, Y., Collangelo, M., Collins, S., Collymore, A., Cook, A.,

Cooke, P., DeArrellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,

Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,

Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,

Hagos, B., Horton, L., Huine, W., Iliev, I., Johnson, R., Jones, C.,

Kamat, A., Karakas, A., Kells, C., Lakocque, K., Lamazares, R.,

Landers, T., Lenocky, J., Levine, R., Liu, G., MacLean, C.,

Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M.,

McEwan, P., McKernan, K., Meldrum, J., Meneus, L., Mihova, T.,

Menga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,

Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J.,

Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C.,

Retts, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J.,

Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S.,

Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,

Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,

Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,

Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,

Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

JOURNAL

Submitted (18-FEB-2002) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 195071)

Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,

Anderson, S., Barna, N., Bastien, V., Boguski, L., Boukhgalter, B.,

Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,

Choquel, Y., Collangelo, M., Collins, S., Collymore, A., Cook, A.,

Cooke, P., DeArrellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,

Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,

Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,

Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,  
 Canarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,  
 Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
 Fato,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,  
 Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,  
 Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,  
 Kamat,A., Karatas,A., Kellis,C., Landers,T., Levine,R.,  
 Lindblad-Toh,K., Liu,G., Maclean,C., Macdonald,P., Major,J.,  
 Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,  
 Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,  
 Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,  
 Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,  
 Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupack,R.,  
 Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,  
 Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,  
 Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,  
 Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission  
 Submitted (31-OCT-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Oct 31, 2002 this sequence version replaced gi:21362181.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.P. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence.submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: L19878  
 Center clone name: 421.C.14  
 ----- Summary Statistics  
 Sequencing vector: Plasmid; n/a; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 193450 bases at least Q40  
 Consensus quality: 193982 bases at least Q30  
 Consensus quality: 194232 bases at least Q20  
 Insert size: 183000; agarose-fp  
 Insert size: 194571; sum-of-contigs  
 Quality coverage: 9.5 in Q20 bases; agarose-fp  
 Quality coverage: 8.9 in Q20 bases; sum-of-contigs

-----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 6 contigs. Gaps between the contigs  
 \* are represented as runs of N. The order of the pieces  
 \* is believed to be correct as given, however the sizes  
 \* of the gaps between them are based on estimates that have  
 \* provided by the submitter.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.  
 \* 1 8394: contig of 8394 bp in length  
 \* 8395 8494: gap of 100 bp  
 \* 8495 9191: contig of 837 bp in length  
 \* 9192 9291: gap of 100 bp  
 \* 9292 121914: contig of 112623 bp in length  
 \* 121915 122014: gap of 100 bp  
 \* 122015 144276: contig of 22262 bp in length  
 \* 144277 144376: gap of 100 bp  
 \* 144377 185733: contig of 41357 bp in length  
 \* 185734 185834: gap of 100 bp  
 \* 185834 195071: contig of 9238 bp in length.

Location/Qualifiers  
 1. 195071  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 /clone="RP23-421C14"  
 /clone\_lib="RPC1-23 Female Mouse BAC"  
 1. 8394  
 /note="assembly\_fragment"

FEATURES  
 source

misc\_feature  
 1. 8394

clone\_end:SP6  
 vector\_side:left  
 8495..19191  
 /note="assembly\_fragment"  
 misc\_feature  
 9292..121914  
 /note="assembly\_fragment"  
 misc\_feature  
 122015..144276  
 /note="assembly\_fragment"  
 misc\_feature  
 144377..185733  
 /note="assembly\_fragment"  
 misc\_feature  
 185834..195071  
 /note="assembly\_fragment"  
 clone\_end:T7  
 vector\_side:right

BASE COUNT 63730 a 34639 c 34804 g 61398 t 500 others  
 ORIGIN

Query Match 18.2%; Score 43.4; DB 2; Length 195071;  
 Best Local Similarity 64.4%; Pred No. 4.5;  
 Matches 65; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 118 AATCTACTGGTTAACTTCAATCAGGAGAAACATGAAGAAGTTAGGAAGTTGATTATCAG 177  
 |||||  
 Db 48286 ATTCTATTTTAAATTTTATATCTTAAAGCCTAAAGCAATGACAAAATTTGTTTATAAG 48345  
 |||||

QY 178 TTGATAGAGTTAGATTCTTCTGGAACCTTTTGAGGGATTCT 218  
 |||||  
 Db 48346 TTATATAGATAGAAATTATGAAGCTTTTGCTAAAAATCT 48386  
 |||||

RESULT 11  
 AC110910

LOCUS AC110910 318620 bp DNA linear HTG 23-APR-2003  
 Mus musculus clone RP24-349E9, WORKING DRAFT SEQUENCE, 23 unordered  
 pieces.

ACCESSION AC110910  
 VERSION AC110910.2 GI:30018095  
 HTG; HTGS PHASE1; HTGS DRAFT.  
 KEYWORDS  
 SOURCE Mus musculus (house mouse)  
 ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
 1 (bases 1 to 318620)  
 Birren,B., Nusbaum,C. and Lander,E.  
 Mus musculus, clone RP24-349E9  
 Unpublished  
 2 (bases 1 to 318620)  
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,  
 Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B.,  
 Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,  
 Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,A.,  
 Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fato,S.,  
 Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S.,  
 Ginde,S., Gord,S., Govette,M., Graham,L., Grand-Pierre,N.,  
 Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,  
 Kamat,A., Karatas,A., Kellis,C., Larocque,K., Lamazares,R.,  
 Landers,T., Lehoczy,J., Levine,R., Liu,G., Maclean,C.,  
 Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,  
 McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,  
 Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,  
 Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,  
 Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,  
 Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,  
 Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R., Seaman,S.,  
 Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
 Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
 Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,  
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission  
 Submitted (16-FEB-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 318620)

TITLE  
 JOURNAL  
 REFERENCE

## AUTHORS

Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, S., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Bouckgaier, B., Canarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., DeArellano, K., Diaz, J.S., Dodge, S., Dooley K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galgan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., MacLean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrim, J., Menes, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, I., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassilev, H., Venkataraman, V.S., Viel, R., VO.A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (23-APR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Apr 17, 2003 this sequence version identified gi:15693473.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
-----Project Information  
Center project name: L21261  
Center clone name: 349 E 9

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 23 'contigs'. The true order of the pieces  
\* is not known and their order in this sequence is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence.  
\* As soon as it is available and the accession number will  
\* be preserved.

1	1220:	contig of 1220 bp in length
1221	1320:	gap of 100 bp
1231	1230:	contig of 1187 bp in length
2508	2607:	gap of 100 bp
2608	4770:	contig of 2163 bp in length
4771	4870:	gap of 100 bp
4871	5338:	contig of 1058 bp in length
5339	6038:	gap of 100 bp
6039	7850:	contig of 1812 bp in length
7851	7950:	gap of 100 bp
7951	10235:	contig of 2285 bp in length
10236	10335:	gap of 100 bp
10336	13293:	contig of 2958 bp in length
13294	13393:	gap of 100 bp
13394	16284:	contig of 2891 bp in length
16285	16384:	gap of 100 bp
16385	19117:	contig of 2733 bp in length
19118	19217:	gap of 100 bp
19218	23060:	contig of 3843 bp in length
23061	23160:	gap of 100 bp
23161	29199:	contig of 6039 bp in length
29200	29299:	gap of 100 bp
29300	36369:	contig of 7670 bp in length
36370	37069:	gap of 100 bp
37070	44337:	contig of 7268 bp in length
44338	44337:	gap of 100 bp
44338	54543:	contig of 10206 bp in length
54544	54743:	gap of 100 bp
54744	64552:	contig of 9509 bp in length

## AUTHORS

```

* 64253 64352: gap of 100 bp
* 64353 76528: contig of 12176 bp in length
* 76529 76628: gap of 100 bp
* 76629 90609: contig of 13981 bp in length
* 90610 90709: gap of 100 bp
* 90710 111194: contig of 20485 bp in length
* 11194 111294: gap of 100 bp
* 111295 126721: contig of 15427 bp in length
* 126722 126722: gap of 100 bp
* 126723 147846: contig of 21025 bp in length
* 147847 147847: gap of 100 bp
* 147848 173784: contig of 25838 bp in length
* 173785 173884: gap of 100 bp
* 173885 217072: contig of 43188 bp in length
* 217073 217172: gap of 100 bp
* 217173 318620: contig of 101448 bp in length.
Location/Qualifiers
1. 318620
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="RP24-349B9"
/clone_lib="RPC1-24 Male Mouse BAC"
1. 1220
/note="assembly_fragment
clone_end:Sp6
vector_side:left"
1321..2507
/note="assembly_fragment"
2608..4770
/note="assembly_fragment"
4871..5938
/note="assembly_fragment"
6039..7850
/note="assembly_fragment"
7951..10235
/note="assembly_fragment"
10336..13293
/note="assembly_fragment"
13394..16284
/note="assembly_fragment"
16385..19117
/note="assembly_fragment"
19218..23060
/note="assembly_fragment"
23161..29199
/note="assembly_fragment"
29300..36969
/note="assembly_fragment"
37070..44337
/note="assembly_fragment"
44438..54643
/note="assembly_fragment"
54744..64252
/note="assembly_fragment"
64353..76528
/note="assembly_fragment"
76629..90609
/note="assembly_fragment"
90710..111194
/note="assembly_fragment"
111295..126721
/note="assembly_fragment"
126822..147846
/note="assembly_fragment"
147947..173784
/note="assembly_fragment"
173885..217072
/note="assembly_fragment"
217173..318620
/note="assembly_fragment
clone_end:T7
vector_side:right"

```

BASE COUNT 97164 a 60953 c 59588 g 98713 t 2202 others  
ORIGIN

Query Match 18.2%; Score 43.4; DB 2; Length 318620;  
Best Local Similarity 64.4%; Pred. No. 4.1;  
Matches 65; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

Qy 118 AATCTACTGGTTAACTTCAATCAGGGGAAACATGACGAAGTTAGGAAGTTGATTATATCAG 177  
|||  
Db 212850 ATTTCTATTTTAAATTTATATCTTAAAGCCTAAGCAATGACAAATTTGTTTTATAAG 212909  
|||  
Qy 178 TTGATAGAGTTAGATTTTCTGGAACTTTTGTGAGGGATTCT 218  
|||  
Db 212910 TTAATATAGATAGAAATATATGAGCTTTTGTGTAATAATTTCT 212950  
|||

RESULT 12  
AX281284 AX281284 11670 bp DNA linear PAT 02-NOV-2001  
DEFINITION Sequence 26 from Patent WO0177164.  
ACCESSION AX281284  
VERSION AX281284.1 GI:16608539  
KEYWORDS synthetic construct  
SOURCE synthetic construct  
ORGANISM artificial sequences.

REFERENCE 1  
AUTHORS Olek.A., Piepenbrock, C. and Berlin, K.  
TITLE Diagnosis of diseases associated with apoptosis  
JOURNAL Patent: WO 0177164-A 26 18-OCT-2001;  
Epigenomics AG, (DE)

FEATURES  
Location/Qualifiers  
1..11670  
/organism="synthetic construct"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32630"  
/note="chemically treated genomic DNA (Homo sapiens)"  
BASE COUNT 3563 a 238 c 2443 g 5426 t  
ORIGIN

Query Match 18.1%; Score 43; DB 6; Length 11670;  
Best Local Similarity 49.3%; Pred. No. 9.4;  
Matches 112; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

Qy 3 TTGGATACGTTGGAAGGCGATATTTGATGGATGGATGGAATGAAGTAAAGCTAGATCA 62  
|||  
Db 265 TTAGGTTATGATTGAAATATATTTTATATATGATTTATATTAATAAATTTTGAAGTGG 324  
|||  
Qy 63 GTTGGGAAAAAGTGTGACACATATCGGTAAAGTGAACCTTATTTGTAATAATCT 122  
|||  
Db 325 GTAAGAATAAATAATGTTTTTATTATAGATTATTAATTTTGTGTTTGGATA 384  
|||  
Qy 123 ACTGGTTAACTCAATCAGGAAACATGAAGTTAGGAAGTTGATTATCAGTTGAT 182  
|||  
Db 385 GAGGTTGATTTAGATCGTAATAATAGTAATAATTAATAATTTAAATTTATGATTAGT 444  
|||  
Qy 183 AGAGTTAGATTTCTGGAACCTTTTGTGAGGATTCATGAACCTG 229  
|||  
Db 445 AGAGTTAGGTTGGTGTGTTTTTATATATTTTATTTTATAGTTTG 491  
|||

RESULT 13  
AL589947/c 143715 bp DNA linear PRI 27-APR-2001  
LOCUS Human DNA sequence from clone RP11-127B16 on chromosome 6, complete  
DEFINITION sequence.  
ACCESSION AL589947 AC064822  
VERSION AL589947.3 GI:13857193  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  
AUTHORS Almeida, J.  
TITLE Direct Submission  
JOURNAL Submitted (27-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
requests: clonerequest@sanger.ac.uk

COMMENT

1 (bases 1 to 143715)  
Almeida, J.  
Direct Submission  
Submitted (27-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
requests: clonerequest@sanger.ac.uk  
On or before May 15, 2001 this sequence version replaced  
gi:7801478, gi:13751578.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest. The following  
abbreviations are used to associate primary accession numbers given  
in the feature table with their source databases: Em: EMBL; Sw:  
SWISSPROT; Tr: TREMBL; Wp: WORMPEP; information on the WORMPEP  
database can be found at  
http://www.sanger.ac.uk/Projects/c\_elegans/wormpep This sequence  
was generated from part of bacterial clone contigs of human  
chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping  
Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr6  
RP11-127B16 is from the library RPCI-11.1 constructed by the group  
of Pieter de Jong. For further details see  
http://www.chori.org/bacpac/home.htm  
VECTOR: pBAC3.6  
IMPORTANT: This sequence is not the entire insert of clone  
RP11-127B16. It may be shorter because we sequence overlapping  
sections only once, except for a 100 base overlap.  
The true left end of clone RP11-538A16 is at 143616 in this  
sequence. The true right end of clone RP11-596A13 is at 100 in this  
sequence.

FEATURES  
source

Location/Qualifiers

1..143715  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/clone="RP11-127B16"  
/clone\_lib="RPCI-11.1"  
954..1833  
/note="L1PA7 repeat: matches 5256..6140 of consensus"  
2918..3386  
/note="L1A7 repeat: matches 1..505 of consensus"  
3803..4812  
/note="pTRS repeat: matches 742..1676 of consensus"  
3817..4406  
/note="CpG island"  
/evidence=not\_experimental  
4813..5485  
/note="L1R12 repeat: matches 1..671 of consensus"  
5486..5500  
/note="pTRS repeat: matches 1676..1690 of consensus"  
5981..6327  
/note="AluY repeat: matches 3..292 of consensus"  
6608..6759  
/note="76 copies 2 mer tt 62% conserved"  
7904..8059  
/note="L1NA5 repeat: matches 6137..6296 of consensus"  
9414..12009  
/note="L1PA6 repeat: matches 3557..6143 of consensus"  
12010..12317  
/note="AluYvA5 repeat: matches 1..308 of consensus"  
12318..12812  
/note="L1PA6 repeat: matches 3067..3557 of consensus"



```
repeat_region 12808..13670
/note="L1Pa6 repeat: matches 2194..3061 of consensus"
repeat_region 15757..15804
/note="24 copies 2 mer ta 75% conserved"
repeat_region 15807..16085
/note="AluSq repeat: matches 4..283 of consensus"
repeat_region 16100..16618
/note="L1ME repeat: matches 5274..5818 of consensus"
repeat_region 16649..16905
/note="AluSq repeat: matches 3..260 of consensus"
repeat_region 16318..17059
/note="L1ME repeat: matches 5100..5255 of consensus"
repeat_region 17302..17674
/note="TH1C repeat: matches 1..371 of consensus"
repeat_region 17675..19219
/note="TH1C-internal repeat: matches 1..1580 of consensus"
repeat_region 19220..19591
/note="TH1C repeat: matches 1..371 of consensus"
repeat_region 19627..19904
/note="L1M4 repeat: matches 4549..4846 of consensus"
repeat_region 19930..20313
/note="TH1C repeat: matches 1..368 of consensus"
repeat_region 21937..22041
/note="FLAM A repeat: matches 42..139 of consensus"
repeat_region 22455..22746
/note="AluSx repeat: matches 1..295 of consensus"
repeat_region 23088..23137
/note="25 copies 2 mer aa 84% conserved"
repeat_region 24198..24565
/note="TH1C repeat: matches 1..371 of consensus"
repeat_region 25952..26154
/note="HA1 repeat: matches 475..784 of consensus"
repeat_region 26159..26513
/note="L2 repeat: matches 1987..2350 of consensus"
repeat_region 27133..27321
/note="WER58A repeat: matches 27..215 of consensus"
repeat_region 27444..27537
/note="HA1 repeat: matches 674..770 of consensus"
repeat_region 27912..28214
/note="L1M4 repeat: matches 1036..1350 of consensus"
repeat_region 29177..33723
/note="L1 repeat: matches 725..5349 of consensus"
repeat_region 37281..37600
/note="WER33 repeat: matches 1..315 of consensus"
repeat_region 38931..39055
/note="L2 repeat: matches 2580..2710 of consensus"
repeat_region 39776..40245
/note="WER1A repeat: matches 60..527 of consensus"
repeat_region 40246..41202
/note="Tigger3b repeat: matches 224..1230 of consensus"
repeat_region 42363..42506
/note="MIR repeat: matches 8..165 of consensus"
repeat_region 43820..43979
/note="MIR repeat: matches 8..157 of consensus"
repeat_region 43980..44493
/note="L1R40a repeat: matches 3..518 of consensus"
repeat_region 44494..44595
/note="MIR repeat: matches 157..261 of consensus"
misc_feature complement(45748..45750)
/note="Sequence from uni-directional primer reads only."
repeat_region 45943..46253
/note="AluSx repeat: matches 1..310 of consensus"
repeat_region 46618..46876
/note="L2 repeat: matches 2433..2744 of consensus"
repeat_region 48916..49048
/note="WER86 repeat: matches 24..153 of consensus"
repeat_region 49255..49644
/note="WSTD repeat: matches 1..394 of consensus"
repeat_region 49645..51316
/note="WSTD-internal repeat: matches 1..1651 of consensus"
repeat_region 51317..51709
/note="WSTD repeat: matches 1..394 of consensus"
```

```
repeat_region 52199..52475
/note="TH1C repeat: matches 94..371 of consensus"
repeat_region 52476..52730
/note="TIGER1 repeat: matches 2078..2338 of consensus"
repeat_region 52774..53152
/note="TIGER1 repeat: matches 1702..2087 of consensus"
repeat_region 53151..53630
/note="TIGER1 repeat: matches 2..484 of consensus"
repeat_region 53953..54262
/note="AluJb repeat: matches 3..303 of consensus"
repeat_region 54608..54785
/note="89 copies 2 mer aa 57% conserved"
repeat_region 54901..55217
/note="L1M4 repeat: matches 2400..2728 of consensus"
repeat_region 55221..55414
/note="WER3 repeat: matches 3..209 of consensus"
repeat_region 55754..55952
/note="L1M1 repeat: matches 5694..5895 of consensus"
repeat_region 56125..56439
/note="AluY repeat: matches 1..311 of consensus"
repeat_region 56685..56980
/note="AluSx repeat: matches 1..296 of consensus"
repeat_region 57012..57287
/note="AluJo repeat: matches 1..285 of consensus"
repeat_region 60017..60050
/note="17 copies 2 mer aa 91% conserved"
repeat_region 61032..61097
/note="33 copies 2 mer ct 71% conserved"
repeat_region 61495..61869
/note="WER47A repeat: matches 1..366 of consensus"
repeat_region 62920..63098
/note="MLTIF repeat: matches 339..540 of consensus"
repeat_region 63117..63220
/note="MLTIC repeat: matches 1..106 of consensus"
repeat_region 63289..63564
/note="MLTIF repeat: matches 29..292 of consensus"
repeat_region 64592..65042
/note="MLTIC repeat: matches 6..394 of consensus"
repeat_region 65287..65563
/note="WSTD repeat: matches 1..394 of consensus"
repeat_region 66238..66349
/note="MIR repeat: matches 28..148 of consensus"
repeat_region 66953..67403
/note="MLTIC repeat: matches 1..460 of consensus"

Query Match 18.1%; Score 43; DB 9; Length 143715;
Best Local Similarity 51.3%; Pred. No. 5.8;
Matches 100; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

QY 43 ATCAAGATAAGCGTAGATGAGTTGGGAAAAAAGTTAGAACATATCGTAAGATGAA 102
Db 29693 ATGTATTATAGCTAATGATTTCGACAGACTCACCAGACATTTATTCGAGAAGAA 29634
QY 103 CGGTATTGTTGAAAAATCTACTGTTACTTCATCAGGGAACATGAAAGTAGG 162
Db 29633 CAGTCTCTTAATAATGTTACCGGAAAAATTTGAATATTATATACAGAAAGAAACT 29574
QY 163 AAGTTGATTATACAGTTGATAGTTAGATTTCCTGGAACTTTTGGAGGGATTCTATGA 222
Db 29573 AGATTACTCTCTACATATATACAAAAATTAACCAAGTGCCTTAAGACTTAATGT 29514
QY 223 AAAACTGGAAGCAAT 237
Db 29513 AAGACCTGAACATAT 29499

RESULT 14
AC021418 161828 bp DNA linear HTG 08-APR-2000
LOCUS Homo sapiens chromosome 4 clone RP11-4E4 map 4, WORKING DRAFT
DEFINITION AC021418
SEQUENCE AC021418 7 unordered pieces.
ACCESSION AC021418
VERSION AC021418.3 GI:7528069
```

```

KEYWORDS HTG; HTGS PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 161828)
JOURNAL Birren, B., Linton, L., Nusbaum, C. and Lander, E.
REFERENCE Homo sapiens Chromosome 4, Clone RP11-4E4
AUTHORS Unpublished
TITLE 2 (bases 1 to 161828)
REFERENCE Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
AUTHORS Anderson, S., Baldwin, J., Barna, N., Becker, R., Beda, F.,
Boguslavskiy, L., Boukhgalter, B., Brown, A., Burkett, G., Castle, A.,
Cooper, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,
DeArelano, K., Dewar, K., Domino, M., Doyle, M., Fenestor, J.,
Ferreira, P., FitzHugh, W., Forrest, C., Gage, D., Galignat, J.,
Gardyna, S., Grant, G., Hages, B., Hearford, A., Horton, L.,
Howland, J., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
Landers, T., Lechoczy, J., Levine, R., Lieu, C., Liu, G., Locke, K.,
Macdonald, P., Marguis, N., McEwan, P., McGurk, A., McKernan, K.,
McPheters, R., Meldrum, J., Meneus, L., Morrow, J., Naylor, J.,
Norman, C., O'Connor, T., O'Donnell, P., Oliver, T., Peterson, K.,
Pierre, N., Pisani, C., Pollard, V., Raymond, C., Riley, R., Rothman, D.,
Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,
Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J.,
Zimmer, A. and Zody, M.
Direct Submission
Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 8, 2000 this sequence version replaced gi:6899731.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WISR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L4929
Center clone name: 4.E.4
----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 157326 bases at least Q40
Consensus quality: 159536 bases at least Q30
Consensus quality: 160510 bases at least Q20
Insert size: 194000; agarose-fp
Insert size: 161228; sum-of-ctngs
Quality coverage: 4.5 in Q20 bases; agarose-fp
Quality coverage: 5.4 in Q20 bases; sum-of-ctngs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 4366: contig of 4366 bp in length
* 4367: gap of 100 bp
* 4467: contig of 11841 bp in length
* 16308: gap of 100 bp
* 16408: contig of 13518 bp in length
* 29925: gap of 100 bp
* 30026: contig of 16799 bp in length
* 46824: gap of 100 bp
* 46825: contig of 16540 bp in length
* 46925: gap of 100 bp
* 63465: contig of 20234 bp in length
* 63466: gap of 100 bp
* 83798: contig of 20234 bp in length

```

```

* 83799 83898: gap of 100 bp
* 83899 161828: contig of 77930 bp in length.
FEATURES
source
1..161828
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone="RP11-4E4"
/clone_lib="RPC1-11 Human Male BAC"
1..4366
misc_feature
4467..16307
/feature="assembly_fragment"
4467..16307
misc_feature
/feature="assembly_fragment"
16408..29925
misc_feature
/feature="assembly_fragment"
clone_end:SP6
vector side:left
30026..46824
misc_feature
/feature="assembly_fragment"
46925..63464
misc_feature
/feature="assembly_fragment"
clone_end:T7
vector side:right
63565..83798
misc_feature
/feature="assembly_fragment"
83899..161828
misc_feature
/feature="assembly_fragment"
BASE COUNT 50687 a 27451 c 28807 g 54283 t 600 others
ORIGIN
Query Match 18.1%; Score 43; DB 2; Length 161828;
Best Local Similarity 51.3%; Pred. No. 5.7;
Matches 100; Conservative 0; Mismatches 95; Indels 0; Gaps 0;
QY 43 ATGAAGATAACCGTAGATGAGTTCGGGAAAAAAGTGTAGAACATATCGGTAGAAATGAA 102
Db 20586 ATGATTATATAGCTAACTGATTTTCGACAGAGTCACCAAGAACATTTATTGGAGAAAGAA 20645
QY 103 CGGTATTGTTGAAAAATCTACTGTTAACTTCAATCAGGAAAAACATGAAGAGTTAGG 162
Db 20646 CAGTCTTCTTAAATGATGGTACCGGAAAAATGATATTTATATACAGAGAGAGAACT 20705
QY 163 AAGTTGATTTATCAGTTGATAGAGTTAGATTTTCTGGAACITTTTGTGAGGATTTCTATGA 222
Db 20706 AGATTACTCTCTCTCACTATATACAAAAATTAACCTCAAGTGGCTTAAGACACTTAAATGT 20765
QY 223 AAAACTGGAGCAAT 237
Db 20766 AAGACCTGAAACTAT 20780
RESULT 15
AL450123
LOCUS Homo sapiens chromosome 6 clone RP11-292M23, *** SEQUENCING IN
DEFINITION PROGRESS ***, 18 unordered pieces.
ACCESSION AL450123
VERSION AL450123.2 GI:11229253
KEYWORDS HTG; HTGS PHASE1; HTGS_CANCELLED.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
REFERENCE Burton, J.
AUTHORS Direct Submission
TITLE Submitted (12-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
JOURNAL CB10 1SA, UK. E-mail enquiries: humquary@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Nov 20, 2000 this sequence version replaced gi:11190634.
----- Genome Center

```

Center: Sanger Centre

Center code: SC  
Web site: <http://www.sanger.ac.uk>  
Contact: [humquery@sanger.ac.uk](mailto:humquery@sanger.ac.uk)

----- Project Information  
Center project name: ba292M23

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Sequencing vector: Plasmid, L08752; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Consensus quality: 194147 bases at least Q40

Consensus quality: 198595 bases at least Q30

Consensus quality: 200843 bases at least Q20

Insert size: 202422; sum-of-contigs

Insert size: 162422; 10.6% error; agarose-fp

Quality coverage: 5.72x in Q20 bases; sum-of-contigs Quality coverage: 8.43x in Q20 bases; agarose-fp

-----  
\* NOTE: This is a 'working draft' sequence. It currently consists of 18 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

* 1 33126: contig of 33126 bp in length
* 33127 33226: gap of 100 bp
* 33227 114642: contig of 81416 bp in length
* 114643 114742: gap of 100 bp
* 114743 123407: contig of 8665 bp in length
* 123408 123507: gap of 100 bp
* 123508 163981: contig of 40474 bp in length
* 163982 164081: gap of 100 bp
* 164082 166908: contig of 2827 bp in length
* 166909 167008: gap of 100 bp
* 167009 169244: contig of 2236 bp in length
* 169245 169344: gap of 100 bp
* 169345 172581: contig of 3237 bp in length
* 172582 172681: gap of 100 bp
* 172682 176172: contig of 3491 bp in length
* 176173 176272: gap of 100 bp
* 176273 179065: contig of 2793 bp in length
* 179066 179165: gap of 100 bp
* 179166 182360: contig of 3195 bp in length
* 182361 182460: gap of 100 bp
* 182461 185537: contig of 3077 bp in length
* 185538 185637: gap of 100 bp
* 185638 188044: contig of 2407 bp in length
* 188045 188144: gap of 100 bp
* 188145 191827: contig of 3683 bp in length
* 191828 191927: gap of 100 bp
* 191928 194758: contig of 2831 bp in length
* 194759 194858: gap of 100 bp
* 194859 197211: contig of 2353 bp in length
* 197212 197311: gap of 100 bp
* 197312 199603: contig of 2292 bp in length
* 199604 199703: gap of 100 bp
* 199704 201839: contig of 2136 bp in length
* 201840 201939: gap of 100 bp
* 201940 204122: contig of 2183 bp in length.
Location/Qualifiers
1...204122
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="6"
/clone="RP11-292M23"
/clone_lib="RPCI-11.2"
1...33126
/note="assembly fragment:01290
fragment_chain:1"
33227...114642

```

#### FEATURES

misc\_feature  
misc\_feature

```

/note="assembly fragment:00700
fragment_chain:1"
114743...123407
/note="assembly fragment:02136
fragment_chain:2"
123508...163981
/note="assembly fragment:03042
fragment_chain:2"
164082...166908
/note="assembly fragment:02528
fragment_chain:3"
167009...169244
/note="assembly fragment:01791
fragment_chain:3"
169345...172581
/note="assembly fragment:00004"
172682...176172
/note="assembly fragment:00332"
176273...179065
/note="assembly fragment:00401"
179166...182360
/note="assembly fragment:00447"
182461...185537
/note="assembly fragment:00980"
185638...188044
/note="assembly fragment:01053"
188145...191827
/note="assembly fragment:02817"
191928...194758
/note="assembly fragment:02913"
194859...197211
/note="assembly fragment:03055"
197312...199603
/note="assembly fragment:03127"
199704...201839
/note="assembly fragment:03282"
201940...204122
/note="assembly fragment:03366"
BASE COUNT 64023 a 36470 c 37269 g 64652 t 1708 others
ORIGIN

```

Query Match 18.1%; Score 43; DB 2; Length 204122;  
Best Local Similarity 51.3%; Pred. No. 5.5;  
Matches 100; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

```

QY 43 ATCAAGATAGCGTAGATGAGTTGGGAAAAAAGTGTAGAACATATCGTAAAGATGAA 102
Db 7545 ATGTTATTTATAGTAACTGATTTTCACACAGCTCCAGAACATTTATTGGAGAAAGAA 7604
QY 103 CCGTTATTTGAAAAAATCTACTGTTTAACTTCAATCAGGGAAAAACATGAGAAAGTTAGG 162
Db 7605 CAGTCTTCTTAATAATGTTACCGGAAAAAATGCAATATTCATATACAGAAAGAAACT 7664
QY 163 AGATTGATTATCAGTTGATAGATTAGTTTCTGGACATTTTCTGAGGGATTCTATCA 222
Db 7665 AGATTACTCTCTCTCACTATATACAAAAAATAACTCAAAAGTGGCTTAAAGACTTAAATGT 7724
QY 223 AAAACTGGAAGCAAT 237
Db 7725 AAGACCTGAAACTAT 7739

```

Search completed: November 15, 2003, 03:34:16  
Job time : 1057.55 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 14, 2003, 23:55:19 ; Search time 153.606 Seconds  
(without alignments)  
4182.570 Million cell updates/sec

Title: US-09-928-457-96

Perfect score: 238

Sequence: 1 AATTGATACGTTGGA...ATGMAAACTGGAGCAATT 238

Scoring table: IDENTITY\_NUC

Gapop 10.0, Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_19Jun03.\*

- 1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*
- 2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*
- 3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*
- 4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*
- 5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*
- 6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*
- 7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*
- 8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*
- 9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*
- 10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*
- 11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*
- 12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*
- 13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*
- 14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*
- 15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*
- 16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*
- 17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*
- 18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*
- 19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*
- 20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*
- 21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*
- 22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*
- 23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*
- 24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*
- 25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	238	100.0	238	19 AAV03604	Neisseria meningitidis
2	233.2	98.0	16677	21 AAA1519	N. meningitidis pa
3	233.2	98.0	349980	21 AAF21612	Neisseria meningitidis
4	233.2	98.0	837096	21 AAA81489	N. meningitidis pa
5	206.8	86.9	1155	21 AAA15302	DNA encoding a pol
6	203.6	85.5	1155	21 AAA15325	DNA encoding a pol
7	200.6	84.3	1152	25 AB238717	N. gonorrhoeae nuc
8	43	18.1	11670	24 ABL54326	Chemically treated

9	40	16.8	20933	24	ABQ67123	Human angiogenesis
10	40	16.8	1664976	19	AAV21209	Methanococcus jann
11	39.8	16.7	6754	24	ABL70346	Chemically treated
12	39.8	16.7	6754	24	AA561305	Human gene regulat
13	39.6	16.6	12592	24	AA561101	Human gene regulat
14	39.2	16.5	6077	24	ABL34218	Human immune syste
15	37.4	15.7	11398	24	ABL34438	Human immune syste
16	36.8	15.5	375	22	ABL25234	Human breast cance
17	36.6	15.4	549	24	ABL78906	Human ovarian can
18	36.4	15.3	7035	22	AA546690	Tumour suppressor
19	36.2	15.2	5928	22	AA546766	Tumour suppressor
20	36.2	15.2	5928	24	ABL34239	Human immune syste
21	36.2	15.2	11260	22	AA54315	Chemically pretrea
22	36.2	15.2	11260	24	ABN80039	Human chemically m
23	36.2	15.2	11260	24	ABK28154	DNA transcription
24	36	15.1	14112	22	AA546478	Tumour suppressor
25	36	15.1	14112	24	ABQ66980	Human angiogenesis
26	36	15.1	14112	24	ABL33443	Human immune syste
27	36	15.1	14112	24	ABK31333	Signal transductio
28	36	15.1	2365589	24	ABA050521	Genomic sequence o
29	35.8	15.0	591	22	AAF82285	Staphylococcus aur
30	35.8	15.0	591	22	AA83221	GTP-binding protei
31	35.8	15.0	617	24	ABK62224	Rat sequence diffe
32	35.8	15.0	1146	24	AA14133	Plasmodium falcipa
33	35.8	15.0	2874	21	AA70144	Plasmodium falcipa
34	35.8	15.0	3107	24	ABK15041	Plasmodium CDNA en
35	35.8	15.0	6124	11	AAQ03568	Sequence encoding
36	35.8	15.0	6124	24	ABK15042	Plasmodium gene fo
37	35.8	15.0	1082138	21	AAF22305	Arabidopsis thalia
38	35.6	15.0	1440	19	AA14319	H. pylori GHPO 128
39	35.6	15.0	123219	23	AAH88703	Human DNA sequence
40	35.4	14.9	1938	20	AAZ221100	Human secreted pro
41	35.4	14.9	6029	24	ABL33992	Human immune syste
42	35.4	14.9	6954	24	ABL33391	Human immune syste
43	35.4	14.9	11398	24	ABL34439	Human immune syste
44	35	14.7	5986	24	AA561432	Human gene regulat
45	35	14.7	5986	24	ABK31498	Signal transductio

ALIGNMENTS

RESULT 1

AAV03604	ID	AAV03604	standard; DNA; 238 BP.
XX	AC	AAV03604;	
XX	DT	22-OCT-1998	(first entry)
XX	DE	Neisseria meningitidis	DNA sequence B87.
XX	KW	N. gonorrhoeae; N. lactamica; chromosome 22491; region 1; region 2;	
XX	KW	region 3; pathogenicity; blood-brain barrier; diagnosis; infection;	
XX	KW	meningitis; ss.	
XX	OS	Neisseria meningitidis.	
XX	PN	WO9802547-A2.	
XX	PD	22-JAN-1998.	
XX	PF	11-JUL-1997;	97WO-FR01295.
XX	PR	12-JUL-1996;	96PR-0008768.
XX	PA	(INRM ) INERM INST NAT SANTE & RECH MEDICALE.	
XX	PA	(PLAC ) MAX PLANCK GSS FOERDERUNG WISSENSCHAFTEN.	
XX	XX	(SMIK ) SMITHKLINE BEECHAM.	
XX	PI	Nassif X, Tinsley C, Achtmann M, Merker P, Ruelle J;	
XX	PI	Vinals C;	



diagnosis; antigen; detection; infection; gene therapy; antibacterial; ds.

Neisseria meningitidis.

WO2000065791-A1.

09-NOV-2000.

08-MAR-2000; 2000WO-US05928.

30-APR-1999; 99US-0132068.

08-OCT-1999; 99WO-US23573.

28-FEB-2000; 2000GB-0004695.

(CHIR ) CHIRON CORP.

(GENO-) INST GENOMIC RES.

Pizza M, Hickey E, Peterson J, Tettelin H, Venter JC, Massignani V; Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V, Rappuoli R; WPI; 2000-647603/62.

Neisseria meningitidis B full length genome sequence and open reading frames are used to detect, treat and prevent Neisserial infections -

Claim 7; Appendix A; 692pp; English.

The present invention describes the full length genome of Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607 to AAF21613 represent fragments of the NMB genomic sequence, as the sequence was too long to go in a record on its own it was split into 8 sequences which overlap each other at the beginning and end of each sequence by 43980 bp (i.e. the last 49980 bp of AAF21544 is repeated at the beginning of AAF21607, the last 49980 bp of AAF21607 are repeated at the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the Neisseria proteins given in AAF58550 to AAF58593, and AAF21589 to AAF21606 represent PCR primers which are used in the exemplification of the present invention. The NMB genome and fragments from it have antibacterial activity, and can be used in vaccines and gene therapy. Neisseria nucleic acids, proteins and/or antibodies which binds to the proteins can be used in compositions for treating or preventing infection due to Neisserial bacteria or as a diagnostic reagent for detecting the presence of Neisserial bacteria or of antibodies raised to Neisserial bacteria. Computers, computer memory, computer storage medium or computer databases can be used in a search to identify open reading frames (ORFs) or coding sequences within the NMB genome. The DNA sequences provide further opportunities to find antigenic or immunogenic proteins which are more effective in vaccines than the outer membrane proteins currently used.

Sequence 349980 BP; 86473 A; 95646 C; 85908 G; 81953 T; 0 other;

Query Match 98.0%; Score 233.2; DB 21; Length 349980;

Best Local Similarity 98.7%; Pred. No. 1.4e-50;

Mismatches 235; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

1 AATTGGATACGTTGGAAAGCGGATATTGATTCGGATGGATGGAAGATGAGT 60

72678 AATTGGATACGTTGGAAAGCGGATATTGATTCGGATGGATGGAAGATGAGT 72619

61 GAGTTGGGAAAAAGTGTAGACATATCGTGAAGATGAACCGTTATTGTTGAAAAAT 120

72618 GAGTTGGGAAAAAGTGTAGACATATCGTGAAGATGAACCGTTATTGTTGAAAAAT 72559

121 CTACTGGTTAACTTCAATCAGGGAACATGACGAAGTTAGGATGATTATCACTTG 180

72558 CTACTGGTTAACTTCAATCAGGGAACATGACGAAGTTAGGATGATTATCACTTG 72499

181 ATAGAGTTAGATTTTCTGGAACTTTGTGGAGGATTTCTATGAAAACTGGAAGCAATT 238

72498 ATAGAGTTAGATTTTCTGGAACTTTGTGGAGGATTTCTATGAAAACTGGAAGCAATT 72441

RESULT 4

AAA81489/C

ID AAA81489 standard; DNA; 837096 BP.

XX

AC AAA81489;

XX

DT 04-DEC-2000 (first entry)

XX

DE N. meningitidis partial DNA sequence gnm\_37 SEQ ID NO:37.

XX

KW Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;

KW antigen; vaccine; diagnosis; infection; antibacterial; identification;

KW Meningococcus B; MenB; ds.

XX

OS Neisseria meningitidis.

XX

PN WO200022430-A2.

XX

PD 20-APR-2000.

XX

PF 08-OCT-1999; 99WO-US23573.

XX

PR 09-OCT-1998; 98US-0103794.

PR 30-APR-1999; 99US-0132068.

XX

PA (CHIR ) CHIRON CORP.

XX

PI Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC; Massignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V; Rappuoli R, Pizza M; WPI; 2000-318079/27.

DR

XX

PT Isolated nucleotide sequences of Neisseria meningitidis which can be used in the diagnosis and treatment of N. meningitidis infection and other Neisserial infections, for example, N.gonorrhoea -

PT

XX

PS Claim 7; Page 629-865; 1760pp; English.

XX

CC The present invention describes methods of obtaining immunogenic proteins from Neisseria genomic sequences. AAA81453 to AAA82414 represent specifically claimed Neisseria meningitidis genomic DNA sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent Neisseria DNA sequences and their corresponding proteins; AAA81254 to AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the isolation of Neisseria meningitidis DNA sequences; and AAA81322 to AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF sequences, which are all used in the exemplification of the present invention. The nucleic acid sequences, protein sequences, and antibodies against them, can be used in the manufacture of a composition. The composition can be used as a medicament (or in the manufacture of a medicament) for treating, preventing or diagnosing infection due to Neisserial bacteria. For example, some of the identified proteins could be components of vaccines against Meningococcus B; against all serotypes; and/or against all pathogenic Neisseriae. Identification of sequences from the bacterium will also facilitate production of biological probes, particularly organism-specific probes. Attempts to make efficacious Meningococcus B vaccines have failed mainly due to antigen tolerance. Multivalent vaccines have also been tried but none have successfully overcome antigenic variability. The provision of further, complete sequences may provide an opportunity to identify secreted or surface exposed proteins that may be presumed targets for the immune system and which are not antigenically variable or at least more conserved than other more variable regions.

XX

SQ Sequence 837096 BP; 207534 A; 227065 C; 205215 G; 197280 T; 2 other;

Query Match 98.0%; Score 233.2; DB 21; Length 837096;

Best Local Similarity 98.7%; Pred. No. 1.7e-50;

Mismatches 235; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AATTGGATACGTTGGAAAAGGGATATTGATTGGGAATGGATGAAGATAAGCGTAGAT 60  
 DB 437449 AATTGGATACGTTGGAAAAGGGATATTGATTGGGAATGGATGAAGATAAGCGTAGAT 437390  
 QY 61 GAGTTGGGAAAAAAGTGTAGAACATATCGTAAAGATGAACCGTTATTGTTGAAAAAT 120  
 DB 437389 GAGTTGGGAAAAAAGTGTAGAACATATCGTAAAGATGAACCGTTATTGTTGAAAAAT 437330  
 QY 121 CTACTGGTTAACTTCAATCAGGAAAAACATGAAGAAGTTAGGAAGTTGATTATCAGTTG 180  
 DB 437329 CTACTGGTTAACTTCAATCAGGAAAAATATGAAGAAGTTAGGAAGTTGATTATCAGTTG 437270  
 QY 191 ATAGAGTTAGATTCTCGAACTTTCTGAGGGATTCTATGAAAACTGGAAGCAATT 238  
 DB 437269 ATAGAGTTAGATTCTCGAAATTTTGTGAGGGATTCTATGAAAACTGGAAGCAATT 437212

## RESULT 5

AAAI5302  
 ID AAI5302 standard; DNA; 1155 BP.

XX  
 AC AAI5302;  
 XX

DT 04-SEP-2000 (first entry)

DE DNA encoding a polypeptide of a *Neisseria* pathogenic strain.

XX Pathogenic strain; *Neisseria*; vaccine; *Neisseria* infection; ss.

OS *Neisseria meningitidis*.

FH Key Location/Qualifiers  
 FT CDS 1..1155  
 FT /\*tag= a

XX WO200026375-A2.  
 PN  
 XX

PD 11-MAY-2000.

PF 28-OCT-1999; 99WO-FR02643.

PR 30-OCT-1998; 98FR-0013693.

XX (INMR ) PASTEUR MERIEUX SERUMS & VACCINS SA.  
 PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.

XX Aujame L, Bouchardon A, Renauld-Mongenie G, Rokbi B, Nassif X;  
 PI Tinsley C, Perrin A;

DR WPI; 2000-365622/31.  
 DR P-PSDB; AAY93271.

XX New polypeptide specific for pathogenic *Neisseria* useful in therapeutic  
 PT or preventative vaccines and for diagnosis

PS Claim 2; Page 73-75; 187pp; French.

XX The present sequence encodes a protein that is specific for pathogenic  
 CC strains of *Neisseria*. The polynucleotides, polypeptides, or their  
 CC antigenic fragments, are used in vaccines to treat or protect against  
 CC *Neisseria* infections, particularly by *N. meningitidis*. The  
 CC polynucleotide sequence is also used for recombinant production of  
 CC the polypeptide and to produce attenuated *Neisseria* strains that  
 CC overexpress it, or express it in a non-toxic mutant form.

XX Sequence 1155 BP; 335 A; 269 C; 248 G; 303 T; 0 other;

Query Match 86.9%; Score 206.8; DB 21; Length 1155;  
 Best Local Similarity 99.0%; Pred. No. 2.9e-44;  
 Matches 208; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AATTGGATACGTTGGAAAAGGGATATTGATTGGGAATGGATGAAGATAAGCGTAGAT 60  
 DB 437449 AATTGGATACGTTGGAAAAGGGATATTGATTGGGAATGGATGAAGATAAGCGTAGAT 437390

DB 946 AATTGGATACGTTGGAAAAGGGATATTGATTGGGAATGGATGAAGATAAGCGTAGAT 1005  
 QY 61 GAGTTGGGAAAAAAGTGTAGAACATATCGTAAAGATGAACCGTTATTGTTGAAAAAT 120  
 DB 1006 GAAATTGGGAAAAAAGTGTAGAACATATCGTAAAGATGAACCGTTATTGTTGAAAAAT 1055  
 QY 121 CTACTGGTTAACTTCAATCAGGAAAAACATGAAGAAGTTAGGAAGTTGATTATCAGTTG 180  
 DB 1066 CTACTGGTTAACTTCAATCAGGAAAAACATGAAGAAGTTAGGAAGTTGATTATCAGTTG 1125  
 QY 181 ATAGAGTTAGATTCTCGAACTTTTGTGCA 210  
 DB 1126 ATAGAGTTAGATTCTCGAACTTTTGTGCA 1155

## RESULT 6

AAAI5325  
 ID AAI5325 standard; DNA; 1155 BP.

XX  
 AC AAI5325;  
 XX

DT 04-SEP-2000 (first entry)

DE DNA encoding a polypeptide of a *Neisseria* pathogenic strain.

XX Pathogenic strain; *Neisseria*; vaccine; *Neisseria* infection; ss.

OS *Neisseria gonorrhoea*.

FH Key Location/Qualifiers  
 FT CDS 1..1155  
 FT /\*tag= a

XX WO200026375-A2.  
 PN  
 XX

PD 11-MAY-2000.

PF 28-OCT-1999; 99WO-FR02643.

PR 30-OCT-1998; 98FR-0013693.

XX (INMR ) PASTEUR MERIEUX SERUMS & VACCINS SA.  
 PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.

XX Aujame L, Bouchardon A, Renauld-Mongenie G, Rokbi B, Nassif X;  
 PI Tinsley C, Perrin A;

DR WPI; 2000-365622/31.  
 DR P-PSDB; AAY93295.

XX New polypeptide specific for pathogenic *Neisseria* useful in therapeutic  
 PT or preventative vaccines and for diagnosis

PS Claim 4; Page 144-146; 187pp; French.

XX The present sequence encodes a protein that is specific for pathogenic  
 CC strains of *Neisseria*. The polynucleotides, polypeptides, or their  
 CC antigenic fragments, are used in vaccines to treat or protect against  
 CC *Neisseria* infections, particularly by *N. meningitidis*. The  
 CC polynucleotide sequence is also used for recombinant production of  
 CC the polypeptide and to produce attenuated *Neisseria* strains that  
 CC overexpress it, or express it in a non-toxic mutant form.

XX Sequence 1155 BP; 337 A; 265 C; 246 G; 307 T; 0 other;

Query Match 85.5%; Score 203.6; DB 21; Length 1155;  
 Best Local Similarity 98.1%; Pred. No. 1.9e-43;  
 Matches 206; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AATTGGATACGTTGGAAAAGGGATATTGATTGGGAATGGATGAAGATAAGCGTAGAT 60  
 DB 946 AATTGGATACGTTGGAAAAGGGATATTGATTGGGAATGGATGAAGATAAGCGTAGAT 1005



QY 61 GAGTTGGGAAAAAGTTAGAACATATCGGTAAAGATGAACCGTTATTGTTGAAAAAT 120  
Db 1006 GAGTTGGGAAAAAGTTAGAACATATCGGTAAAGATGAACCGTTATTGTTGAAAAAT 1065  
QY 121 CTACTGGTTAACTTCAATCAGGAAACATGAAGAGTTAGGAAGTTGATTTATCAGTTG 180  
Db 1066 CTACTGGTTAACTTCAATCAGGAAACATGAAGAGTTAGGAAGTTGATTTATCAGTTG 1125  
QY 181 ATAGAGTTAGATTTCTGGAACCTTTTGCA 210  
Db 1126 ATAGAGTTAGATTTCTGGAATTTTGCA 1155

RESULT 7  
ABZ38717  
ID ABZ38717 standard; DNA; 1152 BP.  
XX AC ABZ38717;  
XX DT 07-MAR-2003 (first entry)  
XX N. gonorrhoeae nucleotide sequence SEQ ID 2023.  
XX DE Antibacterial; infection; vaccine; gene therapy; gene; ds.  
XX KW Neisseria gonorrhoeae.  
XX OS  
XX PN WO200279243-A2.  
XX PD 10-OCT-2002.  
XX PF 12-FEB-2002; 2002WO-IB02069.  
XX PR 12-FEB-2001; 2001GB-0003424.  
XX PA (CHIR-) CHIRON SPA.  
XX PI Fontana MR, Pizza M, Masignani V, Monaci E;  
XX WPI; 2003-058415/05.  
XX DR P-PSDB; ABP77747.  
XX PT New protein from Neisseria gonorrhoeae, useful for the manufacture of a  
XX medicament for treating or preventing N. gonorrhoeae infection -  
XX PS Disclosure; Page 334-335; 815pp; English.  
XX CC The present invention relates to proteins from Neisseria gonorrhoeae.  
CC Also disclosed are the nucleic acid molecules encoding the proteins and  
CC antibodies that specifically bind to the proteins. The composition  
CC comprising the protein nucleic acid or antibody is useful for the  
CC manufacture of a medicament for treating or preventing N. gonorrhoeae  
CC infection, this may be in the form of a vaccine or gene therapy.  
CC Sequences given in records ABZ37706-ABZ42016 represent nucleic acid  
CC molecules of the invention.  
XX SQ Sequence 1152 BP; 336 A; 265 C; 245 G; 306 T; 0 other;

Query Match 84.3%; Score 200.6; DB 25; Length 1152;  
Best Local Similarity 98.1%; Pred. No. 1.2e-42;  
Matches 203; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AATTGGGATACGTTGGAAAGGGATTTGATTGGGAATGGATGAAGATGAAGCGTAGAT 60  
Db 946 AATTGGATACGTTGGAAAGGGATTTGATTGGGAATGGATGAAGATGAAGCGTAGAT 1005  
QY 61 GAGTTGGGAAAAAGTTAGAACATATCGGTAAAGATGAACCGTTATTGTTGAAAAAT 120  
Db 1006 GAGTTGGGAAAAAGTTAGAACATATCGGTAAAGATGAACCGTTATTGTTGAAAAAT 1065  
QY 121 CTACTGGTTAACTTCAATCAGGAAACATGAAGAGTTAGGAAGTTGATTTATCAGTTG 180  
Db 1066 CTACTGGTTAACTTCAATCAGGAAACATGAAGAGTTAGGAAGTTGATTTATCAGTTG 1125

QY 181 ATAGAGTTAGATTTCTGGAACCTTTTG 207  
Db 1126 ATAGAGTTAGATTTCTGGAATTTTG 1152

RESULT 8  
ABL54326  
ID ABL54326 standard; DNA; 11670 BP.  
XX AC ABL54326;  
XX DT 29-JUL-2002 (first entry)  
XX Chemically treated apoptosis gene complementary to gene #13.  
XX DE Apoptosis; HIV; Bloom syndrome; cardiopathy;  
XX KW neurodegenerative disorder; Herpes simplex virus; renal ischaemia;  
XX KW amyotrophic lateral sclerosis; cancer; ds.  
XX OS Unidentified.  
XX PN WO200177164-A2.  
XX PD 18-OCT-2001.  
XX PF 06-APR-2001; 2001WO-EP03969.  
XX PR 06-APR-2000; 2000DE-1019058.  
XX PR 07-APR-2000; 2000DE-1019173.  
XX PR 30-JUN-2000; 2000DE-1032529.  
XX PR 01-SEP-2000; 2000DE-1043826.  
XX PA (EPIG-) EPIGENOMICS AG.  
XX PI Olek A, Piepenbrock C, Berlin K;  
XX WPI; 2002-017444/02.  
XX DR Chemically modified sequences of genes associated with apoptosis are  
XX useful to determine methylation patterns of genomic DNA samples for  
XX diagnosis of associated diseases such as cancer -  
XX PS Claim 1; Seq ID #26; 24pp; English.  
XX CC This invention relates to chemically pre-treated DNA of genes  
CC associated with apoptosis. The nucleic acids are used to allocate  
CC patients for specific therapy for HIV infection, Bloom syndrome,  
CC cardiopathy, aging, neurodegenerative disorders, Herpes simplex virus  
CC infection, renal ischaemia, amyotrophic lateral sclerosis, solid tumours  
CC and cancers. This nucleotide sequence represents a chemically  
CC treated apoptosis gene. Even SEQ ID numbers are the complementary  
CC DNA strands to the odd SEQ ID numbers. The sequence data for this  
CC patent is not represented in the printed specification but is based on  
CC information supplied by the European patent office.  
XX SQ Sequence 11670 BP; 3563 A; 238 C; 2443 G; 5426 T; 0 other;

Query Match 18.1%; Score 43; DB 24; Length 11670;  
Best Local Similarity 49.3%; Pred. No. 0.15;  
Matches 112; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

QY 3 TTGGGATACGTTGGAAAGGGATTTGATTGGGAATGGATGAAGATGAAGCGTAGATGA 62  
Db 265 TTAGGTTAGTTGAAATATATTTTATTTATGGAATATATTTTAAATTTTGAAGTGGG 324  
QY 63 GTTGGGAAAAAGCTTTAGAACATATCGGTAAAGATGAACCGTTATTGTTGAAAAATCT 122  
Db 325 GTAAGAAATAAATGTTTTTTTATTATAGATTATTAATTAATTTTCTTTTGGATA 384  
QY 123 ACTGGTTAACTTCAATCAGGAAACATGAAGAGTTAGGAAGTTGATTTATCAGTTGAT 182  
Db 385 GAGTTGATTTTAGATCGTAAATATAGTAAATATTAATATTTTAAATTTATGATTAGT 444



Db 1095978 AACTAAATTTGCAATACCTGGGGAGGATAGACAAATAGTGGAGATTTATTGTC 1096037

QY 130 AACITCATCAGGGAACATGACAGACTTAGGAGTTGATTATCATCTTATAGAGTTA 189

Db 1096038 TATTCCTAACGAGAAATACAGCAAAATTTGGATTATATGAACCATTAGATAGAGTTA 1096097

QY 190 GATTTCTGGAACTTT 205

Db 1096098 GAGACAGGAGCTTT 1096113

RESULT 11

ID ABL70346

XX ABL70346 standard; DNA; 6754 BP.

AC ABL70346;

XX 01-JUL-2002 (first entry)

DE Chemically treated cell signalling DNA sequence complementary to#118.

XX Cell signalling; cytosine methylation; cell signalling disease;

XX cancer; tumour; cytostatic; ds.

XX Unidentified.

XX WO200202807-A2.

XX 10-JAN-2002.

XX 29-JUN-2001; 2001WO-EP07471.

XX 30-JUN-2000; 2000DE-1032529.

XX 01-SEP-2000; 2000DE-1043826.

XX (EPIG-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPI; 2002-154758/20.

XX Nucleic acid, useful for diagnosis and therapy of diseases associated

PT with cell signalling e.g. cancer, comprises chemically modified genomic

PT sequences of genes associated with cell signalling -

XX Claim 1; SEQ ID NO 236; 24pp+sequence listing; English.

XX The invention relates to a nucleic acid comprising a sequence of at least

CC 18 bases of a segment of chemically pretreated DNA of genes associated

CC with cell signalling. The activity of the modified sequences of the

CC invention may be described as cytostatic. The object of the invention is

CC to provide the chemically modified DNA of genes associated with cell

CC signalling, as well as oligonucleotides and/or PNA-oligomers for

CC detecting cytosine methylations, as well as a method which is

CC particularly suitable for the diagnosis and/or therapy of genetic and

CC epigenetic parameters of genes associated with cell signalling. The

CC chemically modified DNA provided by the invention is useful for diagnosis

CC and therapy of diseases such as solid tumours and cancer. The sequences

CC given in records ABL70111-ABL70626 represent chemically pre-treated

CC genomic DNA's of genes associated with cell signalling.

CC Note: The sequence data for this patent is not represented in the printed

CC specification, but is based on sequence information supplied by the

CC European Patent Office.

XX Sequence 6754 BP; 1864 A; 136 C; 1369 G; 3385 T; 0 other;

SQ

Query Match 16.7%; Score 39.8; DB 24; Length 6754;

Best Local Similarity 50.3%; Pred. No. 0.89;

Matches 98; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

QY 39 TCGGATGAGATACCGTAGTACGCTTCGGGAAACATGTAGAACATATCGGTACAA 98

Db 6028 TAGGAGTCGAGCGGGTGGATTATTGAGGTTAGGATTCGAGATTAGTTGTTAATA 6087

QY 99 TGAACCGTTATTCTTCAAAATCTACTGTTTAACTTCAATCAGGGAAAACATCAAGAAGT 158

Db 6088 TGTGTTTACTTATAGTATAGTGGGGTTTTATTATATATAAAATATATAAAATTTAG 6147

QY 159 TAGGAAGTTTGATTTATCAGTTGATAGAGTTAGATTTTCTGGAACCTTTTGTGAGGGATTT 218

Db 6148 TTAGACGTGTTGATATATATTGTAGTTTATTCGGATTTATTCGGATGTTGAGATAGAGATTT 6207

QY 219 ATGAATAACTGGAG 233

Db 6208 ATTTAAATTTGGGAG 6222

RESULT 12

AAS61305

ID AAS61305 standard; DNA; 6754 BP.

XX AAS61305;

XX 29-JAN-2002 (first entry)

DT Human gene regulation-associated gene oligonucleotide #260.

DE Human; Gene regulation-associated gene; severe combined immunodeficiency;

XX cardiac damage; inflammatory response; Haemophilia; Werner syndrome;

XX asthma; HDR syndrome; congenital heart defect; Saethre-Chotzen syndrome;

XX renal disease; Preeclampsia; cardiac allograft vascular disease;

XX colorectal cancer; thyroid cancer; oesophageal cancer; ds; tumour;

XX immunostimulant; cardiac; antiinflammatory; coagulant; antiasthmatic;

XX nephrotropic; gynecological; anti-tumour; immunosuppressive; cytostatic.

XX Homo sapiens.

OS WO200177375-A2.

XX 18-OCT-2001.

XX 06-APR-2001; 2001WO-EP03968.

XX 06-APR-2000; 2000DE-1019058.

XX 07-APR-2000; 2000DE-1019173.

XX 30-JUN-2000; 2000DE-1032529.

XX 01-SEP-2000; 2000DE-1043826.

XX (EPIG-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPI; 2002-017470/02.

XX New nucleic acid sequences from chemically modified genes associated

PT with gene regulation, useful for analysing cytosine methylations for

PT diagnosis and therapy of diseases e.g. severe combined immunodeficiency

PT disease -

XX Disclosure; SEQ ID No 266; 26pp; English.

XX The invention relates to 224 nucleic acid sequences comprising at least

CC 18 bases of a chemically pretreated gene associated with gene regulation

CC selected from 43 known genes (or complementary sequences). The

CC chemical pretreatment converts cytosine bases unmethylated at the

CC 5-position to uracil or another base with hybridisation behaviour

CC dissimilar to cytosine, to enable analysis of cytosine methylations.

CC The DNA sequences, oligomers (or sets/arrays) and method are

CC useful in the diagnosis of diseases (or predisposition to diseases)

CC associated with gene regulation and in therapy of such diseases, by

CC enabling analysis of the cytosine methylation patterns of such genes,

CC kits are provided. They are especially useful in diagnosis

CC and therapy of e.g. severe combined immunodeficiency disease, cardiac

CC disorders, haemophilia, solid tumours and cancer, Werner syndrome,

CC asthma, HDR syndrome, Saethre-Chotzen syndrome, renal disease,

CC preeclampsia, graft versus-host disease. The present sequence is a

CC sequence included in the sequence data for this specification and is  
 CC associated with the human gene regulation-associated genes.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 6754 BP; 1864 A; 136 C; 1369 G; 3385 T; 0 other;  
 Query Match 16.7%; Score 39.8; DB 24; Length 6754;  
 Best Local Similarity 50.3%; Pred. No. 0.89;  
 Matches 98; Conservative 0; Mismatches 97; Indels 0; Gaps 0;  
 QY 39 TGGGATGAAGATAAGCGTAGATGAGTGGGGAAGAAAGTGTAGAACATATCGGTAGAA 98  
 Db 6028 TAGGAGGTCGAGCGGGTGGATTATTTGAGGTTAGGAGTTCGAGATTAGTTGGTTAATA 6087  
 QY 99 TGAACGGTATTGTGAAATCTACTGCTTAACTTCAATCAGGGAAGAAATCAAGAGAGT 158  
 Db 6088 TGGTGGTTAGTTATAGTATAGTCGGGCTTTTATTTATATATAAAAAATATAAAAAATTAG 6147  
 QY 159 TAGGAGTGTGATTATATCAGTTGATAGATTAGATTTCTGGAATTTTGTGAGGAGTTCT 218  
 Db 6148 TTAGACGGTGGTATATATTTGTTAGTTTATTCGGGATGTTGAGATAGGAGATTT 6207  
 QY 219 ATGAAAGAACTGGAAG 233  
 Db 6208 ATTTAAATTTGGGAG 6222  
 RESULT 13  
 ID AAS61101 standard; DNA; 12592 BP.  
 XX AAS61101;  
 AC AAS61101;  
 XX  
 DT 29-JAN-2002 (first entry)  
 XX  
 DE Human gene regulation-associated gene oligonucleotide #56.  
 KW Human; Gene regulation-associated gene; severe combined immunodeficiency;  
 KW cardiac damage; inflammatory response; Haemophilia; Werner syndrome;  
 KW asthma; HDR syndrome; congenital heart defect; Saethre-Chotzen syndrome;  
 KW renal disease; Freerclampsia; cardiac allograft vascular disease;  
 KW colorectal cancer; thyroid cancer; oesophageal cancer; ds; tumour;  
 KW immunostimulant; cardiast; antiinflammatory; coagulant; antiasthmatic;  
 KW nephrotropic; gynecological; anti-tumour; immunosuppressive; cytostatic.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200177375-A2.  
 XX  
 PD 18-OCT-2001.  
 XX  
 PF 06-APR-2001; 2001WO-EP03968.  
 XX  
 PR 06-APR-2000; 2000DE-1019058.  
 PR 07-APR-2000; 2000DE-1019173.  
 PR 30-JUN-2000; 2000DE-1032529.  
 PR 01-SEP-2000; 2000DE-1043826.  
 XX  
 XX (EPIG-) EPIGENOMICS AG.  
 XX  
 PI Olek A, Piepenbrock C, Berlin K;  
 XX  
 DR WPI; 2002-017470/02.  
 XX  
 XX New nucleic acid sequences from chemically modified genes associated  
 FT with gene regulation, useful for analysing cytosine methylations for  
 PT diagnosis and therapy of diseases e.g. severe combined immunodeficiency  
 PT disease -  
 XX  
 PS Claim 1; SEQ ID No 57; 26pp; English.

XX  
 CC The invention relates to 224 nucleic acid sequences comprising at least  
 CC 18 bases of a chemically pretreated gene associated with gene regulation  
 CC selected from 43 known genes (or complementary sequences). The  
 CC chemical pretreatment converts cytosine bases unmethylated at the  
 CC 5-position to uracil or another base with hybridisation behaviour  
 CC dissimilar to cytosine, to enable analysis of cytosine methylations.  
 CC The DNA sequences, oligomers (or sets/arrays) and method are  
 CC useful in the diagnosis of diseases (or predisposition to diseases)  
 CC associated with gene regulation and in therapy of such diseases, by  
 CC enabling analysis of the cytosine methylation patterns of such genes,  
 CC kits are provided. They are especially useful in diagnosis  
 CC and therapy of e.g. severe combined immunodeficiency disease, cardiac  
 CC disorders, haemophilia, solid tumours and cancer, Werner syndrome,  
 CC asthma, HDR syndrome, Saethre-Chotzen syndrome, renal disease,  
 CC preclampsia, graft versus-host disease. The present sequence is a  
 CC sequence included in the sequence data for this specification and is  
 CC associated with the human gene regulation-associated genes.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 12592 BP; 3783 A; 195 C; 2736 G; 5878 T; 0 other;  
 Query Match 16.6%; Score 39.6; DB 24; Length 12592;  
 Best Local Similarity 48.6%; Pred. No. 1.2;  
 Matches 108; Conservative 0; Mismatches 114; Indels 0; Gaps 0;  
 QY 16 GAAAGGATATTTGATTCGGGAATCGGATGAAGATAGCGTAGATGTCGGGAAAAAA 75  
 Db 7892 GAAGAGATTTTGGTTGGGGAATAATGTTTAGGAGATGATTTATATATGGAATTTAA 7951  
 QY 76 GTGTTAGAACATATCGGTAAGAAATCAACCGTTATTTGTTGAAAAATCTACTGGTTAACTTC 135  
 Db 7952 GAGTAATATATTTGGTTATTTATGAGAAATTAGATTATGCTGTTATGTAATGAAT 8011  
 QY 136 ATCAGGGAACATGAAGAAATGTTAGGAAGTTGATTATTCAGTTGATAGATTGATTT 195  
 Db 8012 AATTATTATATGAGTAGAAGAGTTTTTGGTAGGAGAAATTTATGTAATAAAGATTTATT 8071  
 QY 136 CTGGAACCTTTTGTGAGGAGATTCATGAAAACTGGAAGCAAT 237  
 Db 8072 TCGGAATATTATTGTTGGTTGTTGTTGGAAGGAGAGAGAT 8113  
 RESULT 14  
 ID ABL34218 standard; DNA; 6077 BP.  
 XX ABL34218;  
 AC ABL34218;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Human immune system associated gene SEQ ID NO: 2191.  
 XX  
 KW Human; immune system disease; cytosine methylation; antiasthmatic;  
 KW antiarteriosclerotic; antianaemic; cytostatic; neutropic;  
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;  
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
 KW gene; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200200928-A2.  
 XX  
 PD 03-JAN-2002.  
 XX  
 PF 02-JUL-2001; 2001WO-EP07537.





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 14, 2003, 23:58:39 ; Search time 1587.72 Seconds  
(without alignments)  
3643.257 Million cell updates/sec

Title: US-09-928-457-96

Perfect score: 238

Sequence: 1 AATTGGATACGTTGGAAAA.....ATGAAACTGGAAGCAATT 238

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba.\*

2: em\_esthum.\*

3: em\_estin.\*

4: em\_estmu.\*

5: em\_estov.\*

6: em\_estpl.\*

7: em\_estro.\*

8: em\_htc.\*

9: gb\_est1.\*

10: gb\_est2.\*

11: gb\_hcc.\*

12: gb\_est3.\*

13: gb\_est4.\*

14: gb\_est5.\*

15: em\_estfun.\*

16: em\_estom.\*

17: em\_gss\_hum.\*

18: em\_gss\_inv.\*

19: em\_gss\_pin.\*

20: em\_gss\_vrt.\*

21: em\_gss\_fun.\*

22: em\_gss\_mam.\*

23: em\_gss\_mus.\*

24: em\_gss\_pro.\*

25: em\_gss\_rtd.\*

26: em\_gss\_phg.\*

27: em\_gss\_vrl.\*

28: gb\_gss1.\*

29: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	47.6	20.0	1201	13	BX422549
2	45.8	19.2	1112	9	AL552404
3	44.2	18.6	989	13	BX361014
4	43.2	18.2	905	13	BX369737

C	5	43	18.1	353	28	AQ222887
	6	43	18.1	575	28	AQ506302
	7	43	18.1	1101	29	CNS00067
C	8	42.4	17.8	483	28	B53839
	9	42	17.6	721	14	CB285716
	10	41.8	17.6	1175	13	BX423299
C	11	41.6	17.5	862	28	AZ549439
C	12	41.6	17.5	1101	29	CNS00039G
C	13	41.2	17.3	641	9	AV885336
C	14	41.2	17.3	679	9	AV849915
C	15	41.2	17.3	737	13	BW092427
C	16	41.2	17.3	1083	28	BH180017
	17	41.2	17.3	1083	29	CNS07LYX
C	18	41	17.2	833	13	BQ136570
	19	40.8	17.1	1201	13	BX376686
	20	40.6	17.1	842	29	CNS000KPF
	21	40.4	17.0	1201	13	BX360626
C	22	40.4	17.0	1322	29	CC258640
C	23	40.2	16.9	1191	9	AL558073
	24	40	16.8	519	10	BE432539
C	25	40	16.8	938	29	CNS0006TJ
	26	40	16.8	1101	29	CNS0182P
C	27	39.4	16.6	693	28	BH517222
C	28	39.2	16.5	983	29	CNS014HQ
C	29	39.2	16.5	1201	13	BX356174
	30	39	16.4	434	28	AQ493305
	31	39	16.4	442	13	BQ104067
C	32	38.8	16.3	859	29	CNS0004YY
	33	38.8	16.3	520	13	BQ633195
C	34	38.8	16.3	615	12	BJ405869
C	35	38.8	16.3	855	13	BX355304
C	36	38.8	16.3	1200	13	BX437740
	37	38.8	16.3	1201	13	BX461310
	38	38.6	16.2	855	28	AZ591600
	39	38.6	16.2	861	28	AZ678433
	40	38.6	16.2	971	28	BH165867
	41	38.6	16.2	1101	29	CNS000D1
C	42	38.6	16.2	1101	29	CNS017KX
C	43	38.4	16.1	1201	29	AQ440391
C	44	38.4	16.1	1201	29	CNS01651
C	45	38.2	16.1	737	13	BQ515493

ALIGNMENTS

RESULT 1  
BX422549  
LOCUS  
DEFINITION BX422549 Homo sapiens FETAL LIVER Homo sapiens cDNA clone  
CSODM007YJ05.5-PRIME, mRNA sequence.  
ACCESSION BX422549  
VERSION BX422549.1 GI:30769217  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 1201)  
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
Full-length cDNA libraries and normalization  
Unpublished  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 131 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 8205.f For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CSODM007YJ05.5-PRIME-8205.f.f. Contact :  
Feng Liang Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600

Faraday Avenue Genoscope sequence ID : CSODM007CB03QF1.

## FEATURES

```

source
  Location/Qualifiers
    1. 1201
      /organism="Homo sapiens"
      /mol_type="mRNA"
      /db_xref="taxon:9606"
      /clone="CSODM007YJ05"
      /tissue_type="FETAL LIVER"
      /dev_stage="fetal"
      /clone_lib="Homo sapiens FETAL LIVER"
      /note="Organ: liver; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-strand cDNA was digested with Not I and
cloned into the Not I and EcoRV sites of the pCMVSPORT 6
vector. Library was not normalized."
BASE COUNT      311 a 219 c 275 g 213 t 183 others
ORIGIN
Query Match      20.0%; Score 47.6; DB 13; Length 1201;
Best Local Similarity 45.0%; Pred. No. 0.29;
Matches 86; Conservative 20; Mismatches 85; Indels 0; Gaps 0;

QY 30 GATTGGGAATGGGATGAAGTAAAGCGTAGATGAGTGTGGGAAAAAAGTGTAGAACATAT 89
DB 648 GMTTAAAGAAAGATGAGAGGAGCGGTGSAAGAGSGGGGTGCGAGNATAGARTTAA 707

QY 90 CGGTAAGAAATGAACGGTTATTGTTGAAAAATCTACTCGTTAACTTCAATCAGGGAAACA 149
DB 708 KGTGAAAAAGGAGAAKTTTATGTTATTCGTTTASRAGTTTBGAGAGAGGAGGAAGSAR 767

QY 150 TGAAGAAGTTAGGAGGTTGATTTATCATGTTAGTATGATGTTTCTGGAACTTTGTG 209
DB 768 AGAARAGAGAGGAGCGGTGTTAAATGTGAAAGGGTTTAAATTAAGAAATTTWTTTS 827

QY 210 AGGGAATTCAT 220
DB 828 GGGASSSCTAT 838

RESULT 2
AL552404 1112 bp mRNA linear EST 31-MAY-2003
LOCUS AL552404 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
DEFINITION clone CSODI070YP06 5-PRIME, mRNA sequence.
ACCESSION AL552404
VERSION AL552404.2 GI:31274219
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1112)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
COMMENT On Feb 15, 2001 this sequence version replaced gi:12891270.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 5872.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CSODI070DH03QPI&cluster=5872.r. Contact :
Feng Liang Email : fliang@lifetech.com URL : Corporation 1600
http://fulllength.invitrogen.com/ invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CSODI070DH03QPI.
Location/Qualifiers
  1. 1112
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="taxon:9606"

FEATURES
source
  Location/Qualifiers
    1. 1201
      /organism="Homo sapiens"
      /mol_type="mRNA"
      /db_xref="taxon:9606"

```

```

/clone="CSODI070YP06"
/tissue_type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoRV
sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT      414 a 149 c 189 g 262 t 98 others
ORIGIN
Query Match      19.2%; Score 45.8; DB 9; Length 1112;
Best Local Similarity 38.0%; Pred. No. 0.78;
Matches 90; Conservative 37; Mismatches 110; Indels 0; Gaps 0;

QY 2 ATTGGATACGTGTGAAAGGATATTTCATTGGGAATGGGATGAAGATACCGTAGATG 61
DB 835 MTTTTTTTTTTGTAATAAAWAKAAWTTTAAWTKWTATTTTAAWAGAAARGAARRGGK 894

QY 62 AGTTGGGAAAAAGTGTAGAACATATCGGTAAAGATGAACCGTTATTGTGAAAAATC 121
DB 895 TTTTAAAAAARRKKGAARAAATTTTTTTATKATKATTTTGTKAAAAAATWAA 954

QY 122 TACTGTTAACTTCAATCAGGAAAAACATGAAGAGTTAGGAAGTTGATTTATCAGTTGA 181
DB 955 ATTAGGKAKAWGTTTATGCAAAAAAAGAGAWTTTGTGAATKKAWWATTTTWTWTTT 1014

QY 182 TAGAGTTAGATTTCTCGAACTTTTGTGAGGGATTCATGAAAACTGGAAGCAATT 238
DB 1015 TTAAWAAATTTTIDWWWRGGTTTTTATAAAAKGGGTTTAAAAAATAAATAAW 1071

RESULT 3
BX361014 989 bp mRNA linear EST 05-MAY-2003
LOCUS BX361014 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
DEFINITION clone CSODI078YP09 5-PRIME, mRNA sequence.
ACCESSION BX361014
VERSION BX361014.1 GI:30382486
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 989)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 10967.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CSODI078CH05QPI&cluster=10967.f. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ invitrogen Corporation 1600
Paraday Avenue Genoscope sequence ID : CSODI078CH05QPI.
Location/Qualifiers
  1. 989
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="taxon:9606"
    /clone="CSODI078YP09"
    /tissue_type="PLACENTA COT 25-NORMALIZED"
    /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
    /note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoRV
sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT      358 a 127 c 176 g 152 t 176 others
ORIGIN

```



```

Query Match      18.6%; Score 44.2; DB 13; Length 989;
Best Local Similarity 32.9%; Pred. No. 1.9;
Matches 69; Conservative 52; Mismatches 89; Indels 0; Gaps 0;

QY 4 TTGGATAGCTTGGAAAGGATATTTGATTGGGAATGGATGAACATGAAGCGTAGATGAG 63
Db 686 TRGGGARGKGGKTDAAWAAAWTWTTTKGTGABAAAAAARABAGGGKGTDWAAAA 745

QY 64 TTGGGAAAAAGTGTAGACATATCGGTAGAAGTGAACCGTTATTTGTTGAAAAATCTA 123
Db 746 WTGTGAAAAAAAKTKGGRNRATTAARGGRKWKTKWKTADAADDDKTTAAKATADA 805

QY 124 CTGGTTAACTTCAATCAGGAAAAACATGAAGAAGTTAGGAAGTTGATTATCAGTTGATA 183
Db 806 ADATTTTXXTKAAAAAGGKCAAAWKGGRGAKKATGKTKAAAAAGWTTTXXKTT 865

QY 184 GAGTTAGATTTTCTGGAACCTTTGTGAGGG 213
Db 866 TWTWFAAATTTWKAIAAARTTKRAGARR 895

RESULT 4
LOCUS      BX369737
DEFINITION BX369737 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens
            cDNA clone CS0DC006YG10 5-PRIME, mRNA sequence.
ACCESSION  BX369737
VERSION     BX369737.1
KEYWORDS   EST.
SOURCE     BX369737.1 GI:30453837
           Homo sapiens (human)
ORGANISM   Homo sapiens
REFERENCE  1 (bases 1 to 905)
AUTHORS   Li W.B., Gruber C., Jessee J., and Polayes D.
TITLE     Full-length cDNA libraries and normalization
JOURNAL   Unpublished
COMMENT   Contact: Genoscope
           Genoscope - Centre National de Sequencage
           BP 191 91006 EVRY cedex - France
           Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
           Library was constructed by Life Technologies, a division of
           Invitrogen. This sequence belongs to sequence cluster 3358.f For
           more information about this cluster, see
           http://www.genoscope.cns.fr/
           cgi-bin/cluster.cgi?seq=CS0BAG0052A04_CS00409_1&cluster=3358.f.
           Contact : Feng Liang Email : fliang@lifetech.com URL :
           http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
           Faraday Avenue Genoscope sequence ID : CS0BAG0052A04_CS00409_1.
FEATURES   source
            1..905
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="CS0DC006YG10"
            /tissue_type="NEUROBLASTOMA COT 25-NORMALIZED"
            /clone_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
            /note="1st strand cDNA was primed with a NotI-oligo(dT)
            primer. Five prime end enriched, double-strand cDNA was
            digested with Not I and cloned into the Not I and EcoR V
            sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT  365 a 119 c 163 g 251 t
ORIGIN
Query Match      18.2%; Score 43.2; DB 13; Length 905;
Best Local Similarity 55.9%; Pred. No. 3.2;
Matches 104; Conservative 0; Mismatches 78; Indels 4; Gaps 1;

QY 27 TTTCATTTGGGATCGGATGACATACCGTAGATGAGTTGGGAAAAAAGTGTAGACA 86
Db 381 TTTAATTTGGAAATTTTGGAAAGAACCTTAATTTTGGGGGGGAAAAAATTGAAAAAC 440

```

```

QY 87 TATCGGTAGAAATGAACCGTATTGTTGAAAAATCTACTGGTTAACTCAATCAGGAAA 146
Db 441 TACTAAAAAAATTTAAAGGTTTAAAGGAAAAAATAATTTTAAAGGAAAAAAGGTAA 500

QY 147 ACA----TGAAAGATTTAGGAAGTTGATTTATCAGTTGATAGAGTTAGATTTTCTGGAAC 202
Db 501 AAAAAAGTGAAGAAGCTTGCTGTTGAAAAAGTTTCTGCTGGGTTGATTTATGAAAAT 560

QY 203 TTTTGT 208
Db 561 TTTTIT 566

RESULT 5
LOCUS      AQ222887/c
DEFINITION HS 2013_A1_B04_T7 CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=2013 Col=7 Row=C, genomic survey
            sequence.
ACCESSION  AQ222887
VERSION     AQ222887.1
KEYWORDS   GSS.
SOURCE     AQ222887.1 GI:3636500
           Homo sapiens (human)
ORGANISM   Homo sapiens
REFERENCE  1 (bases 1 to 353)
AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
           Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
           Hood,L.
TITLE     Sequence-tagged connectors: A sequence approach to mapping and
           scanning the human genome
JOURNAL   PROC. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
MEDLINE   99380589
PUBMED    1049764
COMMENT   Contact: Mahairas GG, Wallace JC, Hood L
           High Throughput Sequencing Center
           University of Washington
           401 Queen Anne Avenue North, Seattle, WA 98109, USA
           Tel: (206) 616-3618
           Fax: (206) 616-3887
           Email: jwallace@u.washington.edu
           Sequence Tagged Connector
           Plate: 2013 row: C column: 7
           Class: BAC ends
           High quality sequence stop: 353.
FEATURES   source
            1..353
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /clone="Plate=2013 Col=7 Row=C"
            /sex="male"
            /clone_lib="CIT Approved Human Genomic Sperm Library D"
            /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
            E-Coli DH10B"
BASE COUNT  92 a 94 c 46 g 121 t
ORIGIN
Query Match      18.1%; Score 43; DB 28; Length 353;
Best Local Similarity 53.2%; Pred. No. 2.7;
Matches 91; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 17 AAAAGGGATATTGATTGGGAATGGCATGAAGATAGCGTAGATGAGTTGGGAAAAAAG 76
Db 171 AAAGTAAAAAGTTATATCAGGGTTGGAAGGACACATATACTAAATGTTAGGAGAAAAG 112

QY 77 TCTTAGACATATCGGTAGAATGAACCGTTATTTCTTGAAAAATCTACTGGTTAACTTCA 136
Db 111 TCAATAAATTTTGGTGGACATCCATCAGCGTTAGTAAATATTACTTTTAAATTCG 52

QY 137 ATCAGGAAAAACATGAAGAAGTTAGGAAGTTGATTATCAGTTGATAGAGT 187

```

```
Db 51 GTAGAGGAACCTCAATCAGCAGCAATAGCTCAATTTATAGATGATGATT 1

RESULT 6
AQ506302
LOCUS
DEFINITION
PCPI-11-278E6.TU PCPI-11 Homo sapiens genomic clone PCPI-11-278E6,
genomic survey sequence.
ACCESSION
AQ506302
VERSION
AQ506302.1 GI:4711049
KEYWORDS
GSS.
SOURCE
Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 575)
Zhuo, S., Adams, W.D., Nierman, W., Malek, J., de Jong, P. and Venter
J.C.
TITLE
Use of BAC End Sequences from Library PCPI-11 for Sequence-Ready
Map Building
JOURNAL
Unpublished
COMMENT
Other GSSs: PCPI-11-278E6.TV
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbe@tigr.org
Clones are derived from the human BAC library PCPI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
Research Genet cs (info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: SP6
Class: BAC ends.
FEATURES
Location/Qualifiers
1..575
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/db_xref="GDB:7606469"
/clone_lib="PCPI-11"
/clone="PCPI-11-278E6"
/sex="Male"
/cell_type="Lymphocytes"
/clone_lib="PCPI-11"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
PCPI11 Human Male BAC Library"
BASE COUNT 242 a 99 c 99 g 135 t
ORIGIN
Query Match 18.1%; Score 43; DB 28; Length 575;
Best Local Similarity 51.3%; Pred. No. 3.1;
Matches 100; Conservative 0; Mismatches 95; Indels 0; Gaps 0;
QY 43 ATGAGATAGCGGTAGATGAGTTGGGAAAAAAGTTGTAGAACATATCGGTAAAGATCAA 102
|||||
DB 183 ATGATTTATAGCTAACTGATTTTCGACAGAGTACCAGAACATTATTGGAGAGAGAA 242
QY 103 CGTTTATTGTGAAAAATCTACTGTTAACTTCAATCAGGAAAAACATGAAGAAGTTAGG 162
DB 243 CAGTCTCTTAATAATGTCACCGGAAAAATTTGAATTCATATACAGAAAGAGAAACT 302
QY 163 AGTTGCAITTCAGTTGATGAGTTAGATTTCTCGGAACCTTTGTGAGGATTTCTATGA 222
DB 303 AGATTACTCTCTCTCACTATATACAAAAATTAACATCAAGTGGCTTAAGACTTACATGT 362
QY 223 AAAACTGGAAGCAAT 237
DB 363 AAGACCTGAACTAT 377

RESULT 7
CNS0006J
LOCUS
DEFINITION
Drosophila melanogaster genome survey sequence TET3 end of BAC #
BACR01M22 of PCPI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION
AL062049
VERSION
AL062049.1 GI:4938511
KEYWORDS
GSS.
SOURCE
Drosophila melanogaster (fruit fly)
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 1101)
Direct Submission
Genoscope.
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammosser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named PCPI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
p1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES
Location/Qualifiers
1..1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone_lib="BACR01M22"
/clone="PCPI-98"
/note="end : TET3"
BASE COUNT 333 a 162 c 148 g 177 t 281 others
ORIGIN
Query Match 18.1%; Score 43; DB 29; Length 1101;
Best Local Similarity 24.7%; Pred. No. 3.8;
Matches 55; Conservative 79; Mismatches 89; Indels 0; Gaps 0;
QY 15 GGAAAAGGGATTTTGAATGGGATGGGATGAAGATAGCGGTAGATGTTGGGAAAAA 74
|||||
DB 727 GRANRAAAAKTKKKTKTKRRAAGRRARRRAGAGAAAAAADAATKTGTWAKAA 786
QY 75 AGTTGTAGAACATATCGGTAAAGATGACCGTTATGTTGAAAAATCTACTGTTAATT 134
|||||
DB 787 WRTAAWKDWKATDAAAAATAAARTDWTAKADTKWGAATGAMGAGAAAGRRWGRK 846
QY 135 CAATCAGGAAAAACATGAAGAAGTTAGGAAGTTGATTTTATCATGTTGATAGATTAGTTT 194
|||||
DB 847 GKGDTFARPARAGAGDGDWDKAAWAAWAAATATNWDATATADDDRRKAWDKRAAADR 906
QY 195 TTGGAACCTTTTGTGAGGATTTCTATGAAAAAACTGGAAGCAAT 237
DB 907 DXDKGRAGGTCGWRRRARARTRKAKAKERGAGAKAKKRAK 949

RESULT 8
B53839
LOCUS
DEFINITION
CIT-HSP-2013C7.TF CIT-HSP Homo sapiens genomic clone 2013C7,
genomic survey sequence.
ACCESSION
B53839
```

VERSION	B53839.1	GI:2608173
KEYWORDS	GSS.	
SOURCE	Homo sapiens (human)	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 483)	
AUTHORS	Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and Venter,J.C.	
TITLE	Use of a random BAC End Sequence Database for Sequence-Ready Map Building	
JOURNAL	Unpublished	
COMMENT	Other GSSs: CIT-HSP-2013C7.TR Contact: Mark Adams Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: mdadams@tigr.org Clones are available from Research Genetics (info@resgen.com). BAC end search page: http://www.tigr.org/tcdb/humgen/bac_end_search/bac_end_search.html Seq primer: M13-21 Class: BAC ends.	
FEATURES	Location/Qualifiers	
source	1..483	
	/organism="Homo sapiens"	
	/mol_type="Genomic DNA"	
	/db_xref="GBS:7042233"	
	/db_xref="taxon:9606"	
	/clone="2013C7"	
	/sex="Male"	
	/cell_type="Sperm"	
	/clone_id="CIT-HSP"	
	/note="Vector: pBelOBAcl1; Site_1: HindIII; Site_2: HindIII"	
BASE COUNT	142 a 111 c 75 g 155 t	
ORIGIN		
Query Match	17.8%; Score 42.4; DB 28; Length 483;	
Best Local Similarity	52.9%; Pred. No. 4.1;	
Matches	91; Conservative 0; Mismatches 81; Indels 0; Gaps 0;	
QY	17 AAAAGGATATTGATGGGAATGGGATCAAGATAGCGTAGCTGGGGAAGG 76	
Db	268 AAAGTAAAGTTATATCAGGTTGGAAGAAACAATACTAATGTGTAGGAGAAG 209	
QY	77 TGTAGACATATCGTGAAGATGAACCGTTATGTTGAAAATCTACTGTTAACTTCA 136	
Db	208 TCAATAATTTTGTGGCAATCCATCATCAGCGTTAGTAAATATTACITTTAATCG 149	
QY	137 ATCAGGGAACATGAAGATGGAAGTTGAGTGTGATTTCAGTTCATAGAGTT 188	
Db	148 GTAGAGGACCTCAATCACAGCATCTCTCAATTTATAGATCATGTATT 97	
RESULT 9		
CB285716	721 bp mRNA linear EST 27-FEB-2003	
LOCUS	CMD22_H07_61 UMNPM3 Sus scrofa cDNA clone PPSUBLIB_06H07 5', mRNA	
DEFINITION	sequence.	
ACCESSION	CB285716	
VERSION	CB285716.1	
KEYWORDS	EST.	
SOURCE	Sus scrofa (pig)	
ORGANISM	Sus scrofa	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus. 1 (bases 1 to 721)	
AUTHORS	Dvorak,C.W.T., Hyland,K.A., Zhang,Y., Fahrenkrug,S.C. and Murtaugh,M.P.	

1. 721  
 /organism="Sus scrofa"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9823"  
 /clone="PPSUBLIB\_06407"  
 /tissue\_type="Peyer's patch"  
 /clone\_lib="UNMMPM3"  
 /note="Organ: small intestine, jejunum; Vector: pCMVSPORT6  
 Site 1: NotI; Site 2: SalI; Jejunal Peyer's patches were  
 isolated from healthy, 4-6 week old cross-bred pigs. RNA  
 was extracted either immediately or after 3 hours  
 stimulation in an Using chamber with one of the following  
 treatments: 1) no treatment, 2) salmonella choleraesuis  
 vaccine strain SC-54, 3) phorbol myristate acetate,  
 canavanin A, and 8-bromo-cyclic AMP, and 4)  
 lipopolysaccharide and cholera toxin. Each treatment was  
 performed in the presence and absence of cycloheximide.  
 Purified poly A+ RNA from each of the treated tissues  
 (2-4) was combined together, reverse transcribed, and  
 cloned in to pCMVSPORT6 to make a library of approximately  
 530,000 recombinant clones with an average insert size of  
 1.0 kbp. Poly A+ RNA from freshly isolated and non-treated  
 tissue from an Using chamber (treatment 1) was cloned in  
 the same manner to produce an unstimulated cDNA library of  
 approximately 900,000 clones with an average insert size  
 of 1.5 kbp. Equal portions of the two libraries were  
 pooled and then subtracted with porcine ST fibroblast RNA  
 to create a subtracted porcine Peyer's patch library of  
 approximately 6,000 clones with an average insert size of  
 1.0 kbp."

BASE COUNT 227 a 120 c 124 g 248 t 2 others  
 ORIGIN

Query Match 17.6%; Score 42; DB 14; Length 721;  
 Best Local Similarity 59.0%; Pred. No. 5.8;  
 Matches 72; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 79 TTAGACATATCGGTGAAGTGAACCGTTATTTGGTGAATCTACTGGTTCAAT 138  
 DB 218 TAAACATCTCGGTGAGATCTGCATTTTGTATCTTCAGTCTCGTTAGTAACT 277

QY 139 CAGGGAACATGAGAAAGTTAGGAAGTTGATTTATCAGTTGATGATAGATTTCTG 198  
 DB 278 CAAGGAGAAATAAAGAGTGAATGAATTAATTTTCAATTCGTTAAATTTATG 337

QY 199 GA 200  
 DB 338 AA 339

RESULT 10  
 BX423299  
 LOCUS  
 DEFINITION  
 CS0DF008YB06 3-PRIME, mRNA sequence.  
 ACCESSION  
 BX423299  
 VERSION  
 BX423299.1 GI:30655486  
 EST.  
 KEYWORDS  
 Homo sapiens (human)  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE  
 1 (bases 1 to 1175)  
 Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
 Full-length cDNA libraries and normalization  
 TITLE  
 Unpublished  
 JOURNAL  
 Unpublished  
 COMMENT  
 Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 EVRY cedex - France  
 Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
 Library was constructed by Life Technologies, a division of  
 Invitrogen. This sequence belongs to sequence cluster 8830.r For  
 more information about this cluster see

http://www.genoscope.cns.fr/  
 cgi-bin/cluster.cgi?seq=CSIAF002ZH04NP1&cluster=8830.r. Contact :  
 Feng Liang Email : fliang@lifetech.com URL :  
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
 Faraday Avenue Genoscope sequence ID : CSIAF002ZH04NP1.

FEATURES  
 Location/Qualifiers  
 1..1175  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0DF008YB06"  
 /tissue\_type="FETAL BRAIN"  
 /dev\_stage="fetal"  
 /clone\_lib="Homo sapiens FETAL BRAIN"  
 /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA  
 was primed with a NotI-oligo (dT) primer. Five prime end  
 enriched, double-strand cDNA was digested with Not I and  
 cloned into the Not I and EcoRV sites of the pCMVSPORT 6  
 vector. Library was not normalized."

BASE COUNT 410 a 121 c 115 g 357 t 172 others  
 ORIGIN

Query Match 17.6%; Score 41.8; DB 13; Length 1175;  
 Best Local Similarity 35.5%; Pred. No. 7.5;  
 Matches 75; Conservative 43; Mismatches 93; Indels 0; Gaps 0;

QY 12 GTTGAAAAGGATTTTGTGGGATGGCATGAAGTAAAGCTAGATGAGTTGGGAA 71  
 DB 813 YKKVGTAAAGAGATGGATRAARRAATTGATGATATGGATBEDXTTATTTGRRW 872

QY 72 AAAAGTGTAGACATATCGGTGAAGTGAACCGTTATTTGTGAAAATCTACTGGTTAA 131  
 DB 873 KTTTTRTTTTRATTATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 932

QY 132 CTTCAATCAGGAAACATGAAGACTTAGGAAGTTGATTTATCAGTTAGATGAGTTAGA 191  
 DB 933 TWMRTTTTAMTKRAWARATTTWAAAATAATTTAAATTTTWTWAAAATADTAATTTTATT 992

QY 192 TTTTCTGCAACTTTTGTGAGGATTTCTATGA 222  
 DB 993 TTTTCTTATATTTTATWATAATATWTDW 1023

RESULT 11  
 AZ549439/C  
 LOCUS  
 DEFINITION  
 ENTDP77TF Entamoeba histolytica Sheared DNA linear GSS 14-NOV-2000  
 genomic, genomic survey sequence.  
 ACCESSION  
 AZ549439  
 VERSION  
 AZ549439.1 GI:11174048  
 GSS.  
 KEYWORDS  
 Entamoeba histolytica  
 ORGANISM  
 Entamoeba histolytica  
 Eukaryota; Entamoebidae; Entamoeba.  
 REFERENCE  
 1 (bases 1 to 882)  
 Loftus, B., Van Aken, S. and Fraser, C.  
 Determination of clone end sequences from Entamoeba histolytica  
 TITLE  
 HMI:IMSS sheared DNA library  
 JOURNAL  
 Unpublished  
 COMMENT  
 Contact: Brendan J Loftus  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0208  
 Fax: 301 838 3543  
 Email: bjloftus@tigr.org  
 Clones are derived from the Entamoeba histolytica HMI:IMSS sheared  
 DNA library  
 Seq primer: M13-Forward  
 Class: shotgun  
 High quality sequence start: 23  
 High quality sequence stop: 808.  
 Location/Qualifiers

```

source
1..862
/organism="Entamoeba histolytica"
/mol_type="genomic DNA"
/strain="HM1:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/notes="Vector: pROSL1; Site 1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v+i method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barell, Oxford University Press, 1999)."
BASE COUNT      197 a 156 c 98 g 411 t
ORIGIN
Query Match      17.5%; Score 41.6; DB 28; Length 862;
Best Local Similarity 57.8%; Pred. No. 7.7;
Matches 74; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 37 AATGGGATGAGATGAAGCGTACATGAGTGGGAAAAAGTGTAGACATATCGGTAAG 96
|||||
Db 467 AATGAAGAAGAAAAAGAGTAGAAGAAATAGAGGAAAAAATAATAGAGATATGAAGAG 408
|||||

QY 97 AATGAACCGTTATTGTTGAAAAAATCTACTGTTTCACTTCAATCACTGCGGAAAAACATCAAGAA 156
|||||
Db 407 AAGAAGAAGATGAAGAAAAAGAAACTATTGAATTAATAATAAATAACCAAAATAA 348
|||||

QY 157 GTTAGGAA 164
|||||
Db 347 TTAATGAA 340

RESULT 12
CNS0039G/c.
LOCUS
DEFINITION
Drosophila melanogaster genome survey sequence TET3 end of BAC #
BACR0810 of RPCL-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION      AL063921
VERSION        AL063921.1 GI:4941778
KEYWORDS
SOURCE
ORGANISM
Drosophila melanogaster (fruit fly)
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachyceta; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1101)
Genoscope.
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osogawa and
Aaron Mamoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RCL-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be

```

```

found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES
source
1..1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone_lib="BACR0810"
/clone_lib="RPCL-98"
/notes="end : TET3"
BASE COUNT      201 a 64 c 131 g 202 t 503 others
ORIGIN
Query Match      17.5%; Score 41.6; DB 29; Length 1101;
Best Local Similarity 14.5%; Pred. No. 8.2;
Matches 33; Conservative 128; Mismatches 72; Indels 0; Gaps 0;

QY 2 ATTGATGATGTTGGAAAAAGGATATTGATTGGGAATGGATGAAGAATAAGCGTAGATG 61
|||||
Db 970 DKDKRGDDDKKGGKKKAAKAWATKWDWDDWDDKWDGAKDRKADDDGAGDKD 911
|||||

QY 62 AGTTGGGAAAAAGTGTAGACATATCGTGAAGATGACCGTTATTGTTGAAAAATC 121
|||||
Db 910 DDGKGKDADDGTTGKDDDDDKKWDWDKAKGTWGDATWAAATDWWWGWDADWTTWD 851
|||||

QY 122 TACTGTTTAACTTCAATCAGGAAAAACATGAAGAAGTTAGGAAGTTGATTATCAGTTGA 181
|||||
Db 850 AAADDDWADDWDWAWKNDWAWGARTADRDWDGDRAGKRGARKRDRKADKRD 791
|||||

QY 182 TAGAGTTAGATTCTTCTGGAACCTTTTGTGAGGAGTTCTATGAAAAACTGGAAGCA 235
|||||
Db 790 ADDRDDAATTTTWTTTTTRDTDKWKKTDTWTRWAADRTWDRDDDDDRDRAGTA 737
|||||

RESULT 13
AV885336/c
LOCUS
DEFINITION
AV885336 Nori Satoh unpublished cDNA library, tailbud embryo Ciona
intestinalis cDNA clone rcitb19114 3', mRNA sequence.
ACCESSION      AV885336
VERSION        AV885336.1 GI:16872860
KEYWORDS
SOURCE
ORGANISM
Ciona intestinalis
Ciona intestinalis
Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
Phlebobranchia; Cionidae; Ciona.
REFERENCE
1 (bases 1 to 641)
Sato, N., Satou, Y., Kohara, Y. and Shin-i, T.
Expressed genes in Ciona intestinalis
Unpublished
Contact: Nori Satoh
Department of Zoology
Kyoto University
Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
Tel: 81-75-753-4081
Fax: 81-75-705-1113
Email: sato@ascidian.zool.kyoto-u.ac.jp.
Location/Qualifiers
1..641
/organism="Ciona intestinalis"
/mol_type="mRNA"
/db_xref="taxon:7719"
/clone_lib="rcitb19114"
/tissue_type="whole animal"
/dev_stage="tailbud embryo"
/clone_lib="Nori Satoh unpublished cDNA library, tailbud
embryo"
BASE COUNT      210 a 123 c 107 g 200 t 1 others
ORIGIN
Query Match      17.3%; Score 41.2; DB 9; Length 641;
Best Local Similarity 53.0%; Pred. No. 8.8;
Matches 88; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

```

QY 42 GATGAAGTAAAGCGTAGATAGTTGGGGAAAAAGTGTAGAACATATCGGTAAAGATGA 101  
 |||||  
 Db 605 GCTCAAGTCTCGTAAAAAAGGTTGTCAATAGGAATGTGCACCTTGGCCGAACGGT 546  
 |||||  
 QY 102 ACCGTTATTGTGAAAAATCTACTGGTTAACTTCAATCAGGGAACATGAGAGGTTAG 161  
 |||||  
 Db 545 TGTGTTTTTGTGAAAAATCTACAAATAAGTTGTTATGTTGAGCCCGCCAGGTAAG 486  
 |||||  
 QY 162 GAAGTTGATTTATCAGTTGATAGAGTTAGATTTTCTGGAATTTTG 207  
 |||||  
 Db 485 CCTGTTGATACACAGGTCGTGGAGATGCATTTACTGGAGCTCTTG 440  
 |||||

# RESULT 14 AV849915/c

LOCUS AV849915  
 DEFINITION AV849915 Nori Satoh unpublished cDNA library, cleavage stage embryo  
 Ciona intestinalis cDNA clone rcic104e15 3', mRNA sequence.

ACCESSION AV849915  
 VERSION AV849915.1 GI:16832631

## KEYWORDS SOURCE

ORGANISM Ciona intestinalis

Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;  
 Phlebobranchia; Cionidae; Ciona.

1 (bases 1 to 679)

Satoh,N., Satou,Y., Kohara,Y. and Shin-i,T.

Expressed genes in Ciona intestinalis

UNPUBLISHED

COMMENT

Contact: Nori Satoh

Department of Zoology

Kyoto University

Sakyo-ku, Kyoto, Kyoto 606-8502, Japan

Tel: 81-75-753-4081

Fax: 81-75-705-1113

Email: satoh@ascidian.zool.kyoto-u.ac.jp.

Location/Qualifiers

source

1. 679

/organism="Ciona intestinalis"

/mol\_type="mRNA"

/db\_xref="taxon:7719"

/clone="rcic104e15"

/tissue\_type="whole animal"

/dev\_stage="cleavage stage embryo"

/clone\_lib="Nori Satoh unpublished cDNA library, cleavage stage embryo"

stage embryo"

BASE COUNT 219 a 126 c 109 g 220 t 5 others

ORIGIN

Query Match 17.3%; Score 41.2; DB 9; Length 679;  
 Best Local Similarity 53.0%; Pred.No. 9;  
 Matches 88; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 42 GATGAAGTAAAGCGTAGATAGTTGGGGAAAAAGTGTAGAACATATCGGTAAAGATGA 101  
 |||||  
 Db 641 GCTCAAGTCTCGTAAAAAAGGTTGTCAATAGGAATGTGCACCTTGGCCGAACGGT 582  
 |||||

QY 102 ACCGTTATTGTGAAAAATCTACTGGTTAACTTCAATCAGGGAACATGAGAGGTTAG 161  
 |||||  
 Db 581 TGTGTTTTTGTGAAAAATCTACAAATAAGTTGTTTCATGTTCCAGCCCGCCAGGTAAG 522  
 |||||

QY 162 GAAGTTGATTTATCAGTTGATAGAGTTAGATTTTCTGGAATTTTG 207  
 |||||  
 Db 521 CCTGTTGATACACAGGTCGTGGAGATGCATTTACTGGAGCTCTTG 476  
 |||||

# RESULT 15 BW092427/c

LOCUS BW092427/c  
 DEFINITION BW092427 Nori Satoh unpublished cDNA library, young adult Ciona  
 intestinalis cDNA clone rcic103p01 3', mRNA sequence.

ACCESSION BW092427

VERSION BW092427.1 GI:24268229

## KEYWORDS SOURCE ORGANISM

Ciona intestinalis  
 Ciona intestinalis  
 Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;  
 Phlebobranchia; Cionidae; Ciona.

1 (bases 1 to 737)

Satou,Y., Shin-i,T., Kohara,Y. and Satoh,N.

Expressed genes in Ciona intestinalis (2002c)

UNPUBLISHED

COMMENT

Contact: Nori Satoh

Department of Zoology

Kyoto University

Sakyo-ku, Kyoto, Kyoto 606-8502, Japan

Tel: 81-75-753-4081

Fax: 81-75-705-1113

Email: satoh@ascidian.zool.kyoto-u.ac.jp.

Location/Qualifiers

source

1. 737

/organism="Ciona intestinalis"

/mol\_type="mRNA"

/db\_xref="taxon:7719"

/clone="rcic103p01"

/tissue\_type="whole animal"

/dev\_stage="young adult"

/clone\_lib="Nori Satoh unpublished cDNA library, young adult"

BASE COUNT 236 a 148 c 123 g 230 t

ORIGIN

Query Match 17.3%; Score 41.2; DB 13; Length 737;  
 Best Local Similarity 53.0%; Pred.No. 9.2;  
 Matches 88; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 42 GATGAAGTAAAGCGTAGATAGTTGGGGAAAAAGTGTAGAACATATCGGTAAAGATGA 101  
 |||||  
 Db 617 GCTCAAGTCTCGTAAAAAAGGTTGTCAATAGGAATGTGCACCTTGGCCGAACGGT 558  
 |||||

QY 102 ACCGTTATTGTGAAAAATCTACTGGTTAACTTCAATCAGGGAACATGAGAGGTTAG 161  
 |||||  
 Db 557 TGTGTTTTTGTGAAAAATCTACAAATAAGTTGTTTCATGTTCCAGCCCGCCAGGTAAG 498  
 |||||

QY 162 GAAGTTGATTTATCAGTTGATAGAGTTAGATTTTCTGGAATTTTG 207  
 |||||  
 Db 497 CCTGTTGATACACAGGTCGTGGAGATGCATTTACTGGAGCTCTTG 452  
 |||||

Search completed: November 15, 2003, 08:02:50  
 Job time : 1590.72 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 15, 2003, 00:06:39 ; Search time 37.0704 Seconds  
(without alignments)  
2833.774 Million cell updates/sec

Title: US-09-928-457-96

Perfect score: 238

Sequence: 1 AATTGGATACCTGCGAAAA.....ATGAAAACTCGAAGCAATT 238

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220591566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA: \*  
1: /cgn2\_6/prodata/2/ina/5A-COMB.seq: \*  
2: /cgn2\_6/prodata/2/ina/5B-COMB.seq: \*  
3: /cgn2\_6/prodata/2/ina/5A-COMB.seq: \*  
4: /cgn2\_6/prodata/2/ina/5B-COMB.seq: \*  
5: /cgn2\_6/prodata/2/ina/5A-COMB.seq: \*  
6: /cgn2\_6/prodata/2/ina/5B-COMB.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	41.8	17.6	7218	1	US-08-232-463-14
C 2	40	16.8	1564976	4	US-08-916-421B-1
C 3	38.2	16.1	1287	4	US-09-107-532A-3495
C 4	35.8	15.0	1581	4	US-09-412-778-1
C 5	35.8	15.0	1058	4	US-08-213-419B-20
C 6	35.8	15.0	1074	4	US-08-213-419B-1
C 7	35.8	15.0	6124	4	US-08-213-419B-3
C 8	34.8	14.6	1194	4	US-09-134-COIC-2667
C 9	34.6	14.5	3450	4	US-09-904-615-44
C 10	34.4	14.5	1032	4	US-09-107-532A-2955
C 11	34.2	14.4	1589	3	US-08-747-574-2
C 12	34	14.3	1068	3	US-08-913-842-12
C 13	33	13.9	2981	1	US-08-257-073-2
C 14	33	13.9	2981	2	US-08-184-009-119
C 15	33	13.9	2981	2	US-08-458-356-119
C 16	33	13.9	2981	3	US-08-460-736-119
C 17	33	13.9	2981	4	US-09-535-370-119
C 18	32.8	13.8	3857	3	US-09-347-114A-80
C 19	32.8	13.8	9734	3	US-09-347-114A-80
C 20	32.6	13.7	3001	4	US-09-539-333D-229
C 21	32.2	13.5	2007	4	US-09-328-352-2460
C 22	32	13.4	570	4	US-09-328-352-3114
C 23	32	13.4	1287	4	US-09-134-COIC-964
C 24	32	13.4	1641	4	US-09-601-198-17
C 25	32	13.4	4659	2	US-08-583-276-18
C 26	31.8	13.4	2789	4	US-08-961-527-254
C 27	31.8	13.4	1664976	4	US-08-916-421B-1

28	31.6	13.3	1485	1	US-08-468-036-8	Sequence 8, Appl
29	31.6	13.3	1485	2	US-08-378-843-8	Sequence 8, Appl
30	31.6	13.3	3098	1	US-08-447-500-1	Sequence 1, Appl
31	31.6	13.3	3098	1	US-08-454-097-1	Sequence 1, Appl
32	31.6	13.3	3098	1	US-08-447-408-1	Sequence 1, Appl
33	31.6	13.3	3098	1	US-08-453-866-1	Sequence 1, Appl
34	31.6	13.3	3098	3	US-08-185-359-1	Sequence 1, Appl
35	31.6	13.3	4061	4	US-09-620-312D-363	Sequence 363, App
C 36	31.4	13.2	1830121	4	US-09-557-884-1	Sequence 1, Appl
C 37	31.4	13.2	1830121	4	US-09-643-990A-1	Sequence 1, Appl
38	31	13.0	530	4	US-09-328-352-3314	Sequence 3314, Ap
39	31	13.0	1720	4	US-09-227-357-139	Sequence 139, App
40	31	13.0	2095	4	US-09-227-357-139	Sequence 31, Appl
C 41	31	13.0	2726	1	US-08-461-823-1	Sequence 1, Appl
C 42	31	13.0	3480	1	US-08-920-812-11	Sequence 11, Appl
C 43	31	13.0	3480	1	US-08-920-827-11	Sequence 11, Appl
C 44	31	13.0	3480	1	US-08-921-177-11	Sequence 11, Appl
C 45	31	13.0	3480	1	US-08-362-577C-11	Sequence 11, Appl

ALIGNMENTS

RESULT 1  
US-08-232-463-14/c  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFLINGER, F.  
; APPLICANT: FALKNER, F. G.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 1800 Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22313-0299  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/232,463  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/935,313  
; FILING DATE:  
; APPLICATION NUMBER: EP 91 114 300.6  
; FILING DATE: 26-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 30472/114 INMU  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703)836-9300  
; TELEFAX: (703)683-4109  
; TELEX: 899149  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7218 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; CLONE: pTZgt-F1s  
US-08-232-463-14.

Query Match 17.6%; Score 41.8; DB 1; Length 7218;

```
Best Local Similarity 3.08; Pred. No. 0.02;
Matches 7; Conservative 143; Mismatches 85; Indels 0; Gaps 0;

QY 2 ATTGGATCGTTGGAAAGGATATTGATTGGGAATGGGATGAAGATAGCGTAGATG 61
    ||| ||| : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1444 AATTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1385
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 62 AGTTGGGAAAAGTGTAGACATATCGGTAAGATCAACCGTTATTGTGAAAAATC 121
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1384 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1325
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 122 TACTGTTAACTTCAATCAGGAAACATCAAGAAAGTTAGGAAGTTATTATCAGTTGA 181
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1324 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1265
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 182 TAGAGTAGATTTCTGGAACTTTTGTGAGGATCTATGAATACTGAAGCA 236
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1264 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1210
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 2
US-08-916-421B-1
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
; APPLICANT: Bult et al.
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococcus
; TITLE NO. 6503729
; FILE REFERENCE: PB275
; CURRENT APPLICATION NUMBER: US/08/916,421B
; CURRENT FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/024,428
; PRIOR FILING DATE: 1996-08-22
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 1
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (2822)..(28222)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (28257)..(28258)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84773)..(84773)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84808)..(84808)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84812)..(84812)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98120)..(98120)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98159)..(98159)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98239)..(98239)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98266)..(98266)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98343)..(98343)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (103998)..(103998)
; OTHER INFORMATION: n equals a, t, c, or g

; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (148948)..(148948)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (163385)..(163385)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (191989)..(191989)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (191995)..(191995)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (231980)..(231980)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234187)..(234187)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (779255)..(779255)
; OTHER INFORMATION: n equals a, t, c, or g
```



NAME/KEY: misc feature  
LOCATION: (779576)..(779576)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (855539)..(855539)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (871619)..(871619)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1084830)..(1084830)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1096846)..(1096846)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1113981)..(1113981)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1130881)..(1130881)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1310988)..(1310988)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1313224)..(1313224)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1349473)..(1349473)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1349491)..(1349491)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1470091)..(1470091)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1569020)..(1569020)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1602912)..(1602912)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1603734)..(1603734)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1637998)..(1637998)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1664854)..(1664854)  
OTHER INFORMATION: n equals a, t, c, or g

US-08-916-421B-1

Query Match 16.8%; Score 40; DB 4; Length 1664976;  
Best Local Similarity 55.9%; Pred. No. 0.22;  
Matches 76; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 70 AAAAAGTGTAGAACATATCGGTAGATGAACCGTTATTGTTGAAAATCTACTGTT 129  
DB 1095978 AACTAAATTTTGCATACCTGGGGAGAGATAGCAACAAATAGTGGAGATTTTGGTC 1096037  
QY 130 AACTTCAATCAGGAAAACATGAAGAAGTTAGGAAGTTGATTATCAGTTGATAGAGTTA 189  
DB 1096038 TATTCCTAAGTGAAGAAATACCGCAAAATTTGGATTATATGAACATTAGATAGGTTA 1096097  
QY 190 GATTTTCTGGAACTTT 205  
DB 1096098 CAGACGAGGAGCTTT 1096113

RESULT 3  
US-09-107-532A-3495  
; Sequence 3495, Application US/09107532A

Patent No. 6583275  
GENERAL INFORMATION:  
APPLICANT: Lynn A Doucette-Stamm and David Bush  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
NUMBER OF SEQUENCES: 7310  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
STREET: 100 Beaver Street  
CITY: Waltham  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02354  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: PC  
OPERATING SYSTEM: <Unknown>  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/107,532A  
FILING DATE: 30-Jun-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/085,598  
FILING DATE: 14 May 1998  
APPLICATION NUMBER: 60/051571  
FILING DATE: July 2, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Ariniello, Pamela Denise  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: GTC-012  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-5007  
TELEFAX: (781)893-8277  
INFORMATION FOR SEQ ID NO: 3495:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1287 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: Gouble  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Enterococcus faecium  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...1287  
SEQUENCE DESCRIPTION: SEQ ID NO: 3495:  
US-09-107-532A-3495

Query Match 16.1%; Score 38.2; DB 4; Length 1287;  
Best Local Similarity 56.9%; Pred. No. 0.12;  
Matches 70; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 78 GTTAGACATATCGGTAAAGATGAACCGTTATTGTTGAAAATCTACTGTTAACTCAA 137  
DB 558 GTTGAATAATATCGATAAAAGACTTCTTAATCTTGAAAAATCTCTTAGGAAATGAA 617  
QY 138 TCAGGGAACACATGAAGAAGTTAGGAAGTTGATTATCAGTTGATAGAGTTAGATTCT 197  
DB 618 GTACAGCAGCTATATAAGAGTAGTAAGTAATTTTGTCAATACGATTTTTCAGAAATTC 677  
QY 198 GGA 200  
DB 678 GGA 680

RESULT 4  
US-09-412-779-1  
; Sequence 1, Application US/09412779  
; Patent No. 6352843  
; GENERAL INFORMATION:  
; APPLICANT: Zalacain, Magdalena

```

; APPLICANT: Warren, Richard L.
; TITLE OF INVENTION: ysrc
; FILE REFERENCE: GM10244
; CURRENT APPLICATION NUMBER: US/09/412.779
; CURRENT FILING DATE: 1999-10-04
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 591
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-09-412-779-1

Query Match          15.0%; Score 35.8; DB 4; Length 591;
Best Local Similarity 50.3%; Pred. No. 0.45; 87; Indels 0; Gaps 0;
Matches 88; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

QY 22 GGATATTGATTCGGGAATGGGATCAAGATAGCGTAGATGAGTTGGGGAAAAAAGTGTTA 81
DB 238 GCGTATCGGATATGCTAAAGTAAGTAAACACAAACGTAAGAAAATTTGGAAAAATGATTGAG 297

QY 82 GAACATATCGGTAAAGATGAACCGTTATTGTTGAAAAATCTACTGGTTAACTTCAATCAG 141
DB 298 GAATATATAACTAAGAGAGAGAAATTTGCAATTAGTTATTCAATAGTTGATTTTAAAGACAT 357

QY 142 GGAAAAATCAAGAAGTTAGGAAGTTGATTTATCAGTTGATAGATTAGATTTTC 196
DB 358 GATCCACACAGATGATATCTTAAGTGACAAATTAITGAAATTTGATATTC 412

RESULT 5
US-08-213-419B-20
; Sequence 20, Application US/08213419B
; Patent No. 6333406
; GENERAL INFORMATION:
; APPLICANT: Inselburg, J. et al.
; TITLE OF INVENTION: GENE ENCODING PROTEIN ANTIGENS OF PLASMODIUM FALCIPARUM
; FILE REFERENCE: J11-002CNCIP
; CURRENT APPLICATION NUMBER: US/08/213.419B
; CURRENT FILING DATE: 1994-03-14
; PRIOR APPLICATION NUMBER: US 07/870,506
; PRIOR FILING DATE: 1992-04-17
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
; LENGTH: 1058
; TYPE: DNA
; ORGANISM: Plasmodium falciparum
US-08-213-419B-20

Query Match          15.0%; Score 35.8; DB 4; Length 1058;
Best Local Similarity 59.2%; Pred. No. 0.52;
Matches 61; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 98 ATGAAACCGTTATCTTGGAAAAATCTACTGGTTAACTTCAATCAGGGAACATGAAGAAG 157
DB 20 ATTAAACAAAAATGTTTAAATGTACAGGAGAAAGTCAACACGTTAATACAGGAGGAGG 79

QY 158 TTAGGAAGTTGATTTATCAGTTGATAGATTAGATTTTCTGGA 200
DB 80 TCACGACAGTAAATACACTAGGAGATCAAGCAGGTAGTACAGGA 122

RESULT 6
US-08-213-419B-1
; Sequence 1, Application US/08213419B
; Patent No. 6333406
; GENERAL INFORMATION:
; APPLICANT: Inselburg, J. et al.
; TITLE OF INVENTION: GENE ENCODING PROTEIN ANTIGENS OF PLASMODIUM FALCIPARUM
; FILE REFERENCE: J11-002CNCIP
; CURRENT APPLICATION NUMBER: US/08/213.419B
; CURRENT FILING DATE: 1994-03-14
; PRIOR APPLICATION NUMBER: US 07/870,506
; PRIOR FILING DATE: 1992-04-17
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1058
; TYPE: DNA
; ORGANISM: Plasmodium falciparum
US-08-213-419B-1

```

```

; Sequence 2667, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 2667
; LENGTH: 1194
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-2667

Query Match
Best Local Similarity 14.6%; Score 34.8; DB 4; Length 1194;
Matches 87; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

QY 10 ACCTTGGAAAGGGATATTGTTGGGAATGGGATGAAGATAGCGTAGATGAGTTGGGG 69
Db 179 ACTTAGGTAAGCTTAAAGAGAGAGTGATAAAGAGGACTTTCACTTAAGCCCTGTGCTG 238
QY 70 AAAAAGCTTTAGACATATCGTGAAGATGAACCGTTATTGTTGAAAAAICTACTGTTT 129
Db 239 AAAAATTATCTAAGAAATTAGTAAAGAGATTATTCGTTCTCTGAAACACGTCGCGAAA 298
QY 130 AACTTCAATPACGGGAAACATGAAGAAGTTAGGAAGTTGATTATTCAGTTGATA 183
Db 299 AACTCGAACACGAATTGAACCTTAATGAAGGCGATGTTTATTAGTTGAAA 352

```

```

RESULT 9
US-09-904-615-44/c
; Sequence 44, Application US/09904615
; Patent No. 6566325
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 49 Human Secreted Proteins
; FILE REFERENCE: PZ032P1
; CURRENT APPLICATION NUMBER: US/09/904,615
; CURRENT FILING DATE: 2001-07-16
; PRIOR APPLICATION NUMBER: 09/511,554
; PRIOR FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 60/097,917
; PRIOR FILING DATE: 1998-08-25
; PRIOR APPLICATION NUMBER: 60/098,634
; PRIOR FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 170
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 44
; LENGTH: 3460
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-904-615-44

Query Match
Best Local Similarity 14.5%; Score 34.6; DB 4; Length 3460;
Matches 88; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

QY 1 AATTGGATACGTTGGAAAGGGATATTGTTGGGAATGGGATGAAGATAGCGTAGAT 60
Db 2072 AATATACAAACACTAGGTAATTCAAATTTCTTTGTAATGCGATGGAAGTCTACTGAGAA 2013
QY 61 GAGTTGGGGAAAAAAGCTGTAGACATATCGTGAAGATGAACCGTTTATTGTTGAAAAAT 120
Db 2012 GCAGAAAAAGAACACCATTTTAAACATTATCTGGGCTAGTCAACTGGTAGTGGGTATAAAT 1953

```

```

QY 121 CTACTGTTAACTTCAATCAGGGAAAAACATGAAGAAGTTAGGAAGTTGATTATCAG 177
Db 1952 GTACTGTTAAGGTCAGGCCACACCTAACCTCAATATGAGCCAATTAGTTCACAG 1896

RESULT 10
US-09-107-532A-2955
; Sequence 2955, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 2955:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1092 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...1092
; SEQUENCE DESCRIPTION: SEQ ID NO: 2955:
US-09-107-532A-2955

Query Match
Best Local Similarity 14.5%; Score 34.4; DB 4; Length 1092;
Matches 80; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 31 ATTGGGAATGGGATGAAGATAGCGTAGATGAGTTGGGAAAAAAGTTAGAACATATC 90
Db 737 ATTGTGTGCTAATATGCACACAGCGTGGAGAGATCTAGACAAAAATGATTTAGCAGCTA 796
QY 91 GGTGAAGATGAACCGTTATTGTTGAAAAATCTACTGTTAACTTCATCAGGGAACAT 150
Db 797 GCTTCAAGCAAGTGTGATGATGTTATTGATTACTAAAAACACTTCGAGCTTGTCAAAAT 856
QY 151 GAAGAAGTTAGGAAGTTGATTTATCAGTTGATAGAG 186
Db 857 ATCCAGTCAAGCAACTGGTTGTAGCTGGAGGTAG 892

```



```

; APPLICANT: de Taisne, Charles
; APPLICANT: Time, John A.
; TITLE OF INVENTION: MALARIA RECOMBINANT POXVIRUS VACCINE
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford, P.C.
; STREET: 530 Fifth Avenue, 25th Floor
; CITY: New York
; STATE: New York
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 09-JUN-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/075,783
; FILING DATE: 11-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/852,305
; FILING DATE: 18-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/672,183
; FILING DATE: 20-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Frommer, William S.
; REGISTRATION NUMBER: 25,506
; REFERENCE/DOCKET NUMBER: 454310-2570
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 840-3333
; TELEFAX: (212) 840-0712
; TELEX: 425066 CURTMS
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2981 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; US-08-257-073-2
;
; Query Match 13.9%; Score 33; DB 1; Length 2981;
; Best Local Similarity 60.7%; Pred. No. 3.8;
; Matches 54; Conservative 0; Mismatches 35; Indels 0; Gaps 0;
;
QY 98 ATGAACCGTTATTGTTGAAAAATCTACTGTTAACTTCAATCAGGAGAAAACATGAAGAAG 157
Db 42 ATTACAAAAATGTTATATAATGTACAGGAGAAAGTCAACAGGTAATACAGAGGAGG 101
QY 158 TTAGGAAGTTGATTTATCAGTTGATAGAG 186
Db 102 TCAAGCAGGTAATACAGGAGGAGGTCAAG 130

RESULT 14
US-08-184-009-119
; Sequence 119, Application US/08184009
; Patent No. 5833975
; GENERAL INFORMATION:
; APPLICANT: Paoletti, Enzo
; APPLICANT: Tartaglia, James
; APPLICANT: Cox, William I.
; TITLE OF INVENTION: RECOMBINANT VIRUS IMMUNOTHERAPY
; NUMBER OF SEQUENCES: 217
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford
; STREET: 530 Fifth Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/458,356
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/184,009
; FILING DATE: 19-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Frommer, William S.
; REGISTRATION NUMBER: 25,506

```

```

; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/184,009
; FILING DATE: 19-JAN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Frommer, William S.
; REGISTRATION NUMBER: 25,506
; REFERENCE/DOCKET NUMBER: 454310-2530
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 840-3333
; TELEFAX: (212) 840-0712
; TELEX: 425066 CURTMS
; INFORMATION FOR SEQ ID NO: 119:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2981 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; US-08-184-009-119
;
; Query Match 13.9%; Score 33; DB 2; Length 2981;
; Best Local Similarity 60.7%; Pred. No. 3.8;
; Matches 54; Conservative 0; Mismatches 35; Indels 0; Gaps 0;
;
QY 98 ATGAACCGTTATTGTTGAAAAATCTACTGTTAACTTCAATCAGGAGAAAACATGAAGAAG 157
Db 42 ATTACAAAAATGTTATATAATGTACAGGAGAAAGTCAACAGGTAATACAGAGGAGG 101
QY 158 TTAGGAAGTTGATTTATCAGTTGATAGAG 186
Db 102 TCAAGCAGGTAATACAGGAGGAGGTCAAG 130

RESULT 15
US-08-458-356-119
; Sequence 119, Application US/08458356
; Patent No. 5942235
; GENERAL INFORMATION:
; APPLICANT: Paoletti, Enzo
; APPLICANT: Tartaglia, James
; APPLICANT: Cox, William I.
; TITLE OF INVENTION: RECOMBINANT VIRUS IMMUNOTHERAPY
; NUMBER OF SEQUENCES: 217
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford
; STREET: 530 Fifth Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/458,356
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/184,009
; FILING DATE: 19-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Frommer, William S.
; REGISTRATION NUMBER: 25,506

```

REFERENCE/DOCKET NUMBER: 454310-2530

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 840-3333

TELEFAX: (212) 840-0712

TELEX: 425066CURTMS

INFORMATION FOR SEQ ID NO: 119:

SEQUENCE CHARACTERISTICS:

LENGTH: 2981 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

US-08-458-356-119

Query Match 13.9%; Score 33; DB 2; Length 2981;

Best Local Similarity 60.7%; Pred. No. 3.8;

Matches 54; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

Qy 98 ATGAACCGTTATGTGTAAGAAATCTACTGGTTTCAATCAGGGGAAAACATGAAGAAG 157

Db 42 ATTTACAAAATGTTTATAAATGTACAGGAGAAAGTCAACACAGGTAAATACAGGAGGAGG 101

Qy 158 TTAGGAAGTTGATTTCATCAGTTGATAGAG 186

Db 102 TCAAGCAGGTAATACAGGAGGAGGTCAG 130

Search completed: November 15, 2003, 08:09:42

Job time : 43.0704 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.  
OM nucleic - nucleic search, using sw model  
Run on: November 15, 2003, 00:09:34 ; Search time 134.775 Seconds  
(without alignments)  
5771.357 Million cell updates/sec

Title: US-09-928-457-96  
Perfect score: 238  
Sequence: 1 AATTGGATACGTTGGAAAA.....ATGAAAGCAATTCGACGATTT 238

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2169961 seqs, 1634102185 residues  
Total number of hits satisfying chosen parameters: 4339922

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA:  
1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:  
9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq:  
10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:  
11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:  
12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:  
13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:  
14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:  
15: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:  
16: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:  
17: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	238	100.0	238	10	US-09-928-457-96
2	43	18.1	11670	12	US-10-240-452-26
3	39.2	16.5	6077	12	US-10-311-455-2191
4	38.4	16.1	3673778	12	US-10-312-841-1
5	37.4	15.7	4276	14	US-10-198-846-14002
6	37.4	15.7	11398	12	US-10-311-455-2411
7	37.4	15.7	3673778	12	US-10-312-841-2
8	36.8	15.5	664	12	US-09-814-353-17717
9	36.8	15.5	760	12	US-09-814-353-21458
10	36.8	15.5	760	14	US-10-198-846-13751
11	36.6	15.4	549	10	US-09-867-701-1884
12	36.2	15.2	5928	12	US-10-311-455-2212
13	36.2	15.2	11260	12	US-10-240-453-28
14	36.2	15.2	11260	14	US-10-239-676-20
15	36	15.1	14112	12	US-10-311-455-1416
16	35.8	15.0	617	10	US-09-917-800A-131

17	35.6	15.0	1440	12	US-09-882-227-65	Sequence 65, Appl
18	35.4	14.9	6029	12	US-10-311-455-1965	Sequence 1965, Ap
19	35.4	14.9	6954	12	US-10-311-455-1364	Sequence 1364, Ap
20	35.4	14.9	11398	12	US-10-311-455-2412	Sequence 2412, Ap
21	34.8	14.6	612	12	US-09-814-353-5041	Sequence 5041, Ap
22	34.8	14.6	612	12	US-09-814-353-11333	Sequence 11333, A
23	34.6	14.5	286	10	US-09-878-574-14604	Sequence 14604, A
24	34.6	14.5	3460	9	US-09-739-254-44	Sequence 44, Appl
25	34.6	14.5	3460	9	US-09-904-615-44	Sequence 44, Appl
26	34.6	14.5	3460	12	US-10-055-098-44	Sequence 44, Appl
27	34.6	14.5	3460	14	US-10-054-988-44	Sequence 44, Appl
28	34.6	14.5	6020	12	US-10-204-708-8	Sequence 8, Appl
29	34.6	14.5	6020	12	US-10-311-455-172	Sequence 172, App
30	34.6	14.5	11036	12	US-10-240-453-137	Sequence 137, App
31	34.6	14.5	11036	14	US-10-239-676-117	Sequence 117, App
32	34.4	14.5	5152	12	US-10-204-708-73	Sequence 73, Appl
33	34.4	14.5	6641	12	US-10-311-455-288	Sequence 288, App
34	34.4	14.5	6641	12	US-10-240-453-36	Sequence 36, Appl
35	34.4	14.5	397658	10	US-09-813-320-3	Sequence 3, Appl
36	34.2	14.4	1733	10	US-09-938-842A-3995	Sequence 3995, Ap
37	34.2	14.4	8013	12	US-10-311-455-1480	Sequence 1480, Ap
38	34.2	14.4	18598	12	US-10-311-455-360	Sequence 360, App
39	34	14.3	1135	12	US-10-027-632-119273	Sequence 119273,
40	34	14.3	1135	13	US-10-027-632-119273	Sequence 119273,
41	34	14.3	6917	12	US-10-311-455-2208	Sequence 2208, Ap
42	34	14.3	8530	12	US-10-311-455-408	Sequence 408, App
43	34	14.3	10133	12	US-10-311-455-432	Sequence 432, App
44	33.8	14.2	5396	12	US-10-240-483-153	Sequence 153, App
45	33.8	14.2	16724	12	US-10-311-455-1064	Sequence 1064, Ap

ALIGNMENTS

RESULT 1

US-09-928-457-96  
; Sequence 96, Application US/09928457  
; Patent No. US20020164603A1  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: DNA, specific proteins and peptides  
; TITLE OF INVENTION: of the Neisseria meningitidis species bacteria, method  
; TITLE OF INVENTION: for obtaining them and their biological application.  
; NUMBER OF SEQUENCES: 99  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30 (OEB)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/928,457  
; FILING DATE: 2001-08-14  
; PRIOR APPLICATION DATA: 09/214,759  
; FILING DATE: 199-12-10  
; INFORMATION FOR SEQ ID NO: 96:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 238 base pairs  
; TYPE: nucleotide  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHEetical: NO  
; ANTI-SENSE: NO  
; US-09-928-457-96

Query Match 100.0%; Score 238; DB 10; Length 238;  
Best Local Similarity 100.0%; Pred. No. 5.9e-51;  
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AATTGGATACGTTGGAAAGGATATTTGATTCGGATGGATGAAGATAGCGTAGAT 60  
|||||  
Db 1 AATTGGATACGTTGGAAAGGATATTTGATTCGGATGGATGAAGATAGCGTAGAT 60  
|||||

```
QY 61 GAGTTGGGAAAAAGTGTAGACATATCGTAAAGATGAACCGTTATTGTTGAAAAAT 120
Db 61 GAGTTGGGAAAAAGTGTAGACATATCGTAAAGATGAACCGTTATTGTTGAAAAAT 120
QY 121 CTACTGGTTAACTTCAATCAGGAAAAACATGAAGAAGTTAGGAAGTTGATTTATCAGTTG 180
Db 121 CTACTGGTTAACTTCAATCAGGAAAAACATGAAGAAGTTAGGAAGTTGATTTATCAGTTG 180
QY 181 ATAGAGTTAGATTTTCTGGAACCTTTTGTGAGGATTTCTATCAAAAACCTGGAACCAAT 238
Db 181 ATAGAGTTAGATTTTCTGGAACCTTTTGTGAGGATTTCTATCAAAAACCTGGAACCAAT 238

RESULT 2
US-10-240-452-26
; Sequence 26, Application US/10240452
; Publication No. US20030162194A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Apoptosis
; FILE REFERENCE: 5013.1006
; CURRENT APPLICATION NUMBER: US/10/240.452
; PRIOR FILING DATE: 2002-10-02
; PRIOR APPLICATION NUMBER: PCT/EP01/03969
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 78
; SEQ ID NO 26
; LENGTH: 11670
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-452-26

Query Match 18.1%; Score 43; DB 12; Length 11670;
Best Local Similarity 49.3%; Pred. No. 0.8;
Matches 112; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

QY 3 TTTCGATACGTTGGAAAAAGGATATTTCGATTGGGAATCGGATGAAGATAACCGTAGATGA 62
Db 265 TTAGGTATGATGTAAGAAATATATTTTATATGGAATTATATTAATAAATTTGAAGTGGG 324
QY 63 GTTGGGAAAAAGTGTAGACATATCGTAAAGATGAACCGTTATTGTTGAAAAATCT 122
Db 325 GTAAGAAATATAAATGTTTTTATATAGATTATTAATAAATTTTGTGTTTGGATA 384
QY 123 ACTGTTAACTTCAATCAGGAAAAACATGAAGAAGTTAGGAAGTTGATTTATCAGTTGAT 182
Db 385 GAGGTGATTTTATAGATCGTAAATAATAGTAATTAATAATTTTAAATTTAGTTAGT 444
QY 183 AGAGTTAGATTTTCTGGAACCTTTTGTGAGGATTTCTATCAAAAACCTG 229
Db 445 AGAGTTAGCTGCTGCTGCTTTTATATATTTTATATTTTATATGTTG 491

RESULT 3
US-10-311-455-2191
; Sequence 2191, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
```

```
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determi
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311.455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 2191
; LENGTH: 6077
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-2191

Query Match 16.5%; Score 39.2; DB 12; Length 6077;
Best Local Similarity 50.5%; Pred. No. 5.8;
Matches 95; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

QY 24 ATATTGATTGGGAATGGCATGAAGATGAACGTTAGATGAGTTGGGAAAAAGTGTAGA 83
Db 824 ATAGTGGTTTGGGGTGGAAAGTTTTTAATGGAATTGTTGGTGTGTTGTTGGATGA 883
QY 84 ACATATCGGTAAGAATGAACCGTTATTGTTGAAAAAATCTACTGTTAACTTCAATCAGGG 143
Db 884 GTATTTAATTTTGGTTTGGAGTTTTTTGAAGTTTTTTTGGTTAAGATGATTTCAG 943
QY 144 AAAACATGAAGAAGTTAGGAAGTTGATTTATCAGTTAGATGATTTTCGGAAT 203
Db 944 GAAATGATTGATTATTGTTGTTGTTTTTTAAATCGATGATATAGTTTAGATTTTATAAT 1003
QY 204 TTTGTGAG 211
Db 1004 TATATGAG 1011

RESULT 4
US-10-312-841-1
; Sequence 1, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
; FILE REFERENCE: B01/1208/WO
; CURRENT APPLICATION NUMBER: US/10/312.841
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 1
; LENGTH: 367378
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; NAME/KEY: unsure
; LOCATION: (3294164)
US-10-312-841-1

Query Match 16.1%; Score 38.4; DB 12; Length 367378;
Best Local Similarity 49.5%; Pred. No. 93;
Matches 99; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

QY 25 TATTGATTGGGAATCGGATGAAGATGAACGTTAGATGAGTTGGGAAAAAGTGTAGA 84
Db 433638 TTTTATTTTATTTAGTTTGAATAAGTAGGTGTGTTGTTTTTTAGGAACGATTTAATCGTC 433697
QY 85 CATATCGGTAAGAATGAACCGTTATTGTTGAAAAAATCTACTGTTAACTTCAATCAGGGA 144
```





```

; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR FILING DATE: US 60/191,031
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17717
; LENGTH: 664
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-814-353-17717

Query Match      15.5%; Score 36.8; DB 12; Length 664;
Best Local Similarity 54.4%; Pred. No. 10;
Matches 74; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 95 AGAATGAACCGTTATTGTTGAAAATCTACTGTTTCAATCAGGGAACATGAAG 154
Db 627 ACAATCCAGTTTGTGAGAAAGAAACAATTCATCTATATCTGCAAAATGTAAT 568
QY 155 AAGTTAGGAAGTTGATTTATCAGTTGATAGAGTTTCTGGAACCTTTGTGAGGGA 214
Db 567 ATTTAGGAATAGTCTCTAGTTAATAATGTTTACAGTTCTAATCATCTGTTTGA 508
QY 215 TTCTATGAAAACCTGG 230
Db 507 TTTGTGAGAAATGG 492

RESULT 9
US-09-814-353-21458/c
; Sequence 21458, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21

```

```

; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21458
; LENGTH: 760
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; LOCATION: 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168,
; LOCATION: 169, 170, 171, 172, 173, 174, 175
; OTHER INFORMATION: n = A,T,C or G
; US-09-814-353-21458

Query Match      15.5%; Score 36.8; DB 12; Length 760;
Best Local Similarity 54.4%; Pred. No. 11;
Matches 74; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 95 AGAATGAACCGTTATTGTTGAAAATCTACTGTTTCAATCAGGGAACATGAAG 154
Db 522 ACAATCCAGTTTGTGAGAAAGAAACAATTCATCTATATCTGCAAAATGTAAT 463
QY 155 AAGTTAGGAAGTTGATTTATCAGTTGATAGAGTTTCTGGAACCTTTGTGAGGGA 214
Db 462 ATTTAGGAATAGTCTCTAGTTAATAATGTTTACAGTTCTAATCATCTGTTTGA 403
QY 215 TTCTATGAAAACCTGG 230
Db 402 TTTGTGAGAAATGG 387

RESULT 10
US-10-198-846-13751/c
; Sequence 13751, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13751
; LENGTH: 760
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-198-846-13751

Query Match      15.5%; Score 36.8; DB 14; Length 760;
Best Local Similarity 54.4%; Pred. No. 11;
Matches 74; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 95 AGAATGAACCGTTATTGTTGAAAATCTACTGTTTCAATCAGGGAACATGAAG 154
Db 522 ACAATCCAGTTTGTGAGAAAGAAACAATTCATCTATATCTGCAAAATGTAAT 463
QY 155 AAGTTAGGAAGTTGATTTATCAGTTGATAGAGTTTCTGGAACCTTTGTGAGGGA 214
Db 462 ATTTAGGAATAGTCTCTAGTTAATAATGTTTACAGTTCTAATCATCTGTTTGA 403
QY 215 TTCTATGAAAACCTGG 230
Db 402 TTTGTGAGAAATGG 387

RESULT 11

```

US-09-867-701-1884/c

; Sequence 1884, Application US/09867701  
; Patent No. US20020132237A1  
; GENERAL INFORMATION:  
; APPLICANT: Aglate, Paul A.  
; APPLICANT: Harlocker, Susan L.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER  
; FILE REFERENCE: 21021.497  
; CURRENT APPLICATION NUMBER: US/09/867,701  
; CURRENT FILING DATE: 2001-05-29  
; NUMBER OF SEQ ID NOS: 10912  
; SOFTWARE: Fast-Seq for Windows Version 4.0  
; SEQ ID NO 1884  
; LENGTH: 549  
; TYPE: DNA  
; ORGANISM: Homo sapien  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(549)  
; OTHER INFORMATION: n = A,T,C or G

US-09-867-701-1884

Query Match 15.4%; Score 36.6; DB 10; Length 549;  
Best Local Similarity 55.2%; Pred. No. 11;  
Matches 69; Conservative 0; Mismatches 56; Indels 0; Gaps 0;  
QY 106 TTATTCTGAAATCTACTGGTTAACTCAATCAGGGAAACATGAAGAGTTAGGAAG 165  
Db 335 TTGTGAGAAGAACTAATCACTCTATCTGCAATGTAAATNNTAGGAA 276  
QY 166 TTGATTATCAGTTGATAGATAGATTTCTGGAACCTTTGTGAGGATCTATGAAAA 225  
Db 275 TTAGTCTAGGTTAATAATGTTTACAGTCTTAACATCTGCTTTTAGATTTCGTGAAA 216  
QY 226 ACTGG 230  
Db 215 ATGG 211

RESULT 12

US-10-311-455-2212  
; Sequence 2212, Application US/10311455  
; Publication No. US20030143606A1

; GENERAL INFORMATION:  
; APPLICANT: OLEK, Alexander  
; APPLICANT: PIEPENBROCK, Christian  
; APPLICANT: BERLIN, Kurt  
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining Methylation of Cytosine  
; TITLE OF INVENTION: cytosine methylation  
; FILE REFERENCE: 5013.1014  
; CURRENT APPLICATION NUMBER: US/10/311,455  
; CURRENT FILING DATE: 2002-12-16  
; PRIOR APPLICATION NUMBER: PCT/EP01/07537  
; PRIOR FILING DATE: 2001-07-02  
; PRIOR APPLICATION NUMBER: DE 10032529.7  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: DE 10043826.1  
; PRIOR FILING DATE: 2000-09-01  
; NUMBER OF SEQ ID NOS: 2424  
; SEQ ID NO 2212  
; LENGTH: 5928  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

US-10-311-455-2212

Query Match 15.2%; Score 36.2; DB 12; Length 5928;  
Best Local Similarity 53.1%; Pred. No. 33;  
Matches 77; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 25 TATTGATTGGGAATGGGATGAAGATAACCGTAGATGAGTTGGGAAAAAAGTTAGAA 84  
Db 5179 TATTGATTGGGAATGGGATGAAGATAACCGTAGATGAGTTGGGAAAAAAGTTAGAA 5238  
QY 85 CATATCGGTAAAGAATGACCGTTATTGTTGAAAAATCTACTGTTAACTTCAATCAGGA 144  
Db 5239 GAATATTTTCATTAAGAATGAGTATTGAGTAATAATTATTAGATTATTTTCAAGTTGGT 5298  
QY 145 AACATGAAGAAGTTAGGAAGTTGA 169  
Db 5299 GAAAGCTATAAAATTTATAAAATTTA 5323

RESULT 13

US-10-240-453-28  
; Sequence 28, Application US/10240453  
; Publication No. US20030148326A1  
; GENERAL INFORMATION:  
; APPLICANT: OLEK, Alexander  
; APPLICANT: PIEPENBROCK, Christian  
; APPLICANT: BERLIN, Kurt  
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA  
; TITLE OF INVENTION: Transcription  
; TITLE OF INVENTION: By Means of Assessing the Methylation Status of Genes Associated with DNA  
; TITLE OF INVENTION: With DNA Transcription  
; FILE REFERENCE: 5013.1009  
; CURRENT APPLICATION NUMBER: US/10/240,453  
; CURRENT FILING DATE: 2002-10-02  
; PRIOR APPLICATION NUMBER: PCT/EP01/03973  
; PRIOR FILING DATE: 2001-04-06  
; PRIOR APPLICATION NUMBER: DE 10019058.8  
; PRIOR FILING DATE: 2000-04-06  
; PRIOR APPLICATION NUMBER: DE 10019173.8  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: DE 10032529.7  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: DE 10043826.1  
; PRIOR FILING DATE: 2000-09-01  
; NUMBER OF SEQ ID NOS: 350  
; SEQ ID NO 28  
; LENGTH: 11260  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

US-10-240-453-28

Query Match 15.2%; Score 36.2; DB 12; Length 11260;  
Best Local Similarity 48.3%; Pred. No. 42;  
Matches 101; Conservative 0; Mismatches 108; Indels 0; Gaps 0;  
QY 18 AAAGGGATATTTGATTGGGAATGGGATGAAGATAACCGTAGATGAGTTGGGAAAAAAGT 77  
Db 9770 AAAGTCTTATATTAATGCTGTGTAAGAGTAGAATAATAATAATTTTTCAGAAAAA 9829  
QY 78 GTTAGAACATATCGGTGAAGAATGAACCGTTATTGTTGAAAAATCTACTGTTAACTCAA 137  
Db 9830 ATAATAAAATATATTAGGAAATAGGAATAAATTTAAGTAATAATAATTTTAAATTTTA 9889  
QY 138 TCAGGGAACATGAGGAGTTAGGAGTTAGTATTCAGTTGATAGATTAGATTTTCT 197  
Db 9890 TAGAGTAATTAATGAGTAGTTATTGTTAGATATATATAAGAAAAAGTGTATTTT 9949  
QY 198 GGAACATTTTGTGAGGAGTTCTATGAAAAA 226  
Db 9950 GGAATCTTTTATATGTTTGTGGGTTAA 9978

RESULT 14

US-10-239-676-20  
; Sequence 20, Application US/10239676  
; Publication No. US20030082609A1  
; GENERAL INFORMATION:

APPLICANT: OLEK, Alexander  
APPLICANT: PIEPENBROCK, Christian  
TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation  
FILE REFERENCE: 5013.1003  
CURRENT APPLICATION NUMBER: US/10/239,676  
CURRENT FILING DATE: 2002-09-24  
PRIOR APPLICATION NUMBER: PCT/EP01/03968  
DE 10019058.8  
DE 10019173.8  
DE 10032529.7  
DE 10043826.1  
PRIOR FILING DATE: 2001-04-06  
2000-04-06  
2000-04-07  
2000-06-30  
2000-09-01  
NUMBER OF SEQ ID NOS: 228  
SEQ ID NO 20  
LENGTH: 11260  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-239-676-20

Query Match 15.2%; Score 36.2; DB 14; Length 11260;  
Best Local Similarity 48.3%; Pred. No. 42;  
Matches 101; Conservative 0; Mismatches 108; Indels 0; Gaps 0;  
QY 18 AAGGATATTTGATTTGGGAATGGGATGAAGATAGCGGTAGATGATGCGGAAAAAAGT 77  
DB 9770 AAAGTGTATATATATGCTGTGAAAAAGTAGAATAATATATTTTTTTTTCAGAAAAA 9829  
QY 78 GTTAGAACATATCGGTAAAGATGAACCGTTATTGTGAAAAATCTACTGTTTAACTCAA 137  
DB 9830 ATATTAATATATATAGGAATAGGAATAATTAAGTAATAATTTTAAATTTTA 9889  
QY 138 TCAGGAAACATGAAGAAGTTAGGAAGTTGATTTATCAGTTGATAGATTTCT 197  
DB 9890 TAGAGGTAATTTATGAAGTAGTTATGCTAGATATATATAAGAAAAAGTGTTTTATTTT 9949  
QY 198 GGAACCTTTTGTAGGCGATCTATGAAAA 226  
DB 9950 GGAATGTTATATGTTTGTGGGTTAA 9978

RESULT 15  
US-10-311-455-1416  
Sequence 1416, Application US/10311455  
Publication No. US20030143606A1  
GENERAL INFORMATION:  
APPLICANT: OLEK, Alexander  
APPLICANT: PIEPENBROCK, Christian  
TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Peter  
FILE REFERENCE: 5013.1014  
CURRENT APPLICATION NUMBER: US/10/311,455  
CURRENT FILING DATE: 2002-12-16  
PRIOR APPLICATION NUMBER: PCT/EP01/07537  
PRIOR FILING DATE: 2001-07-02  
PRIOR APPLICATION NUMBER: DE 10032529.7  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: DE 10043826.1  
PRIOR FILING DATE: 2000-09-01  
NUMBER OF SEQ ID NOS: 2424  
SEQ ID NO 1416  
LENGTH: 14112  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

US-10-311-455-1416

Query Match 15.1%; Score 36; DB 12; Length 14112;  
Best Local Similarity 49.0%; Pred. No. 51;  
Matches 96; Conservative 0; Mismatches 100; Indels 0; Gaps 0;  
QY 13 TTGGAAAAGGGATATTTGATTTGGGAATGGGATGAAGATAGCGGTAGATGATGCGGAAAA 72  
DB 1492 TTGAATATGGAGTTTATATAATTTTCGTTTAAAGCTTTTAAGGGTTGTTGGGTAATTTGAGGAAA 1551  
QY 73 AAAGTGTTAGAACATATCGGTAAAGATGAACCGTTATTGTGAAAAATCTACTGTTTAAAC 132  
DB 1552 AAAGGTGCTAGGTTTGGTTGAGNAATTAGGGTTTTTTTTTAATATTCGGGTAGTAGG 1611  
QY 133 TTCAATCAGGGAACATGAAGAAGTTAGGAAGTTGATTTATCAGTTGATAGATTTAGAT 192  
DB 1612 GGTTCGTACGTGTAAAGATAGAGAGGGTATGGGTTTCGTTGTTTTTTTATTTTCGTTTTAT 1671  
QY 193 TTTCTGGAACTTTTGT 208  
DB 1672 TTTTTCGGGTTTGT 1687

Search completed: November 15, 2003, 09:32:46  
Job time : 145.875 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 14, 2003, 23:56:24 ; Search time 960.589 Seconds  
(without alignments)  
9241.609 Million cell updates/sec

Title: US-09-928-457-99

Perfect score: 217

Sequence: 1 AATTCGGACAGTATGAAATAC.....ATGACGATGAGATTAAATT 217

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq.length: 0

Maximum DB seq.length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl :

1: gb.ba.\*

2: gb.htg.\*

3: gb.in.\*

4: gb.om.\*

5: gb.ov.\*

6: gb.pat.\*

7: gb.ph.\*

8: gb.pl.\*

9: gb.pr.\*

10: gb.ro.\*

11: gb.sts.\*

12: gb.sy.\*

13: gb.un.\*

14: gb.vi.\*

15: em.ba.\*

16: em.fun.\*

17: em.hum.\*

18: em.in.\*

19: em.mu.\*

20: em.om.\*

21: em.or.\*

22: em.ov.\*

23: em.pat.\*

24: em.ph.\*

25: em.pl.\*

26: em.ro.\*

27: em.sts.\*

28: em.un.\*

29: em.vi.\*

30: em.htg.hum.\*

31: em.htg.inv.\*

32: em.htg.other.\*

33: em.htg.mus.\*

34: em.htg.pln.\*

35: em.htg.rod.\*

36: em.htg.vrt.\*

37: em.htg.vrt.\*

38: em.sy.\*

39: em.htgo.hum.\*

40: em.htgo.mus.\*

41: em.htgo.other.\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	217	100.0	217	1	AF169477
2	217	100.0	217	6	A68928
3	217	100.0	217	6	BD063028
4	191.4	88.2	349061	1	NW0222491
5	130.4	60.1	858	6	AX024058
6	130.4	60.1	858	6	AX024111
7	42.8	19.7	223954	2	AC106472
8	42.8	19.7	261007	2	AC133419
9	41.4	19.1	163284	10	AC129204
10	41.4	19.1	183202	10	AC125488
11	40.4	18.6	181815	2	AC131594
12	40.4	18.4	301450	1	AP003188
13	39.2	18.1	695	3	AF412737
14	39.2	18.1	695	3	AF412737
15	39.2	18.1	163453	2	EX088572
16	38.8	17.9	153336	2	AC025020
17	38.8	17.9	156277	9	AC107015
18	38.8	17.9	163022	2	AC016851
19	38.8	17.9	163737	9	AC108697
20	37.6	17.3	602	6	AX381792
21	37.6	17.3	3507	9	AF077367
22	37.6	17.3	3574	9	BC018122
23	37.6	17.3	3615	9	D88687
24	37.6	17.3	3817	6	BD094040
25	37.6	17.3	3826	9	AR274914
26	37.6	17.3	3826	9	HSTHIRED
27	37.6	17.3	3826	9	S79851
28	37.6	17.3	83755	9	AC090107
29	37.6	17.3	114103	5	AL672171
30	37.6	17.3	251845	2	AC097949
31	37.6	17.3	303250	1	AF003193
32	37.4	17.2	173929	9	AC007217
33	37.4	17.1	38432	8	SPAC11E3
34	36.8	17.0	235700	2	AC103577
35	36.6	16.9	177987	9	AC107052
36	36.2	16.7	155568	2	AP005393
37	36.2	16.7	238256	2	AC094282
38	36.2	16.7	242935	2	AC094306
39	36.2	16.6	2000	6	AX509017
40	36	16.6	2003	6	AX461324
41	36	16.6	2004	6	AX461324
42	36	16.6	3390	9	HSTHIREDU
43	36	16.6	102873	8	ATF26GS
44	36	16.6	110000	6	BD061520_5
45	36	16.6	110000	6	BD061520_5

ALIGNMENTS

RESULT 1  
AF169477  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE

AF169477 217 bp DNA linear BCT 09-AUG-2000  
Neisseria meningitidis strain 22491 clone Em10 unknown sequence.  
AF169477  
AF169477.1 GI:9754685  
Neisseria meningitidis  
Neisseria meningitidis  
Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
Neisseriaceae; Neisseria.  
1 (bases 1 to 217)  
Perrin,A., Nassif,X. and Tinsley,C.R.  
Identification of regions of the chromosome of Neisseria  
meningitidis and Neisseria gonorrhoeae which are specific to

Pred. No. is the number of results predicted by chance to have a

```

pathogenic Neisseriae
Unpublished
REFERENCE 2 (bases 1 to 217)
AUTHORS Perrin,A., Nassif,X. and Tinsley,C.R.
TITLE Direct Submission
JOURNAL Submitted (16-JUL-1999) Necker Medicine Faculty, INSERM U411, 156
rue de Vaugivard, Paris 75015, France
FEATURES
    source
        Location/Qualifiers
            1..217
                /organism="Neisseria meningitidis"
                /mol_type="genomic DNA"
                /strain="Z2491"
                /db_xref="taxon:487"
                /clone="Em10"
BASE COUNT 76 a 31 c 44 g 56 t
ORIGIN
    Query Match 100.0%; Score 217; DB 1; Length 217;
    Best Local Similarity 100.0%; Pred. No. 8.1e-44;
    Matches 217; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AATTCGGACAGTATGAATACAGCGGATTAATACAGGTAAGTTTCATTACACGGAAGAAC 60
DB 1 AATTCGGACAGTATGAATACAGCGGATTAATACAGGTAAGTTTCATTACACGGAAGAAC 60
QY 61 CTTTAAAGAAATAATGAAGGTATTACCTTGTTGCCAACGGGAATGCTAAATATGCC 120
DB 61 CTTTAAAGAAATAATGAAGGTATTACCTTGTTGCCAACGGGAATGCTAAATATGCC 120
QY 121 GAGTTTTCTACTGAATACAGCAATCCAGCAATTCATTTCATATTTGACTGATGCTGAA 180
DB 121 GAGTTTTCTACTGAATACAGCAATCCAGCAATTCATTTCATATTTGACTGATGCTGAA 180
QY 181 TGTGGACTTTATAGATAATGACGATGAAGATTAAAT 217
DB 181 TGTGGACTTTATAGATAATGACGATGAAGATTAAAT 217
RESULT 2
BD063028 217 bp DNA linear PAT 06-MAY-1999
LOCUS
DEFINITION
    DNA and specific proteins or peptides of the NEISSERIA MENINGITIDIS
    species bacteria, method for obtaining them and their biological
    applications.
ACCESSION
    BD063028
VERSION
    BD063028.1 GI:22608631
KEYWORDS
    JP 2001504684-A/90.
SOURCE
    unidentified
ORGANISM
    unclassified.
REFERENCE
    1 (bases 1 to 217)
AUTHORS Nassif,X., Tinsley,C., Achtman,M., Ruelle,J.L., Vinals,C. and
    Merker,P.
TITLE
    DNA and specific proteins or peptides of the Neisseria meningitidis
    species bacteria, method for obtaining them and their biological
    applications
JOURNAL
    Patent: JP 2001504684-A 90 10-APR-2001;
    INSTITUT NATIONAL DE LA SANTE ET DE LA RECHERCHE MEDICALE, MAX
    PLANCK GESELLSCHAFT ZUR FORDERUNG DER WISSENSCHAFTEN EV BERLIN,
    SMITHKLINE BEECHAM
COMMENT
    PN JP 2001504684-A/90
    PD 10-APR-2001
    PF 11-JUL-1997 JP 1998505685
    PR 12-JUL-1996 FR 96/08768
    PI XAVIER NASSIF, COLIN TINSLEY, MARK ACHTMAN, JEAN LOUIS RUELLE, PI
    CARLA VINALS.
    PI PETRA MERKER
    PC C12N15/31,C07K14/22,C07K16/12,A61K39/095,C12Q1/68,G01N33/53 CC
    Strandedness: Single;
    CC Topology: Linear;
    FH Key Location/Qualifiers.
FEATURES
    source
        1..217
            /organism="unidentified"
            /mol_type="genomic DNA"
            /db_xref="taxon:32644"
BASE COUNT 76 a 31 c 44 g 66 t
ORIGIN
    Query Match 100.0%; Score 217; DB 6; Length 217;
    Best Local Similarity 100.0%; Pred. No. 8.1e-44;
    Matches 217; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AATTCGGACAGTATGAATACAGCGGATTAATACAGGTAAGTTTCATTACACGGAAGAAC 60
DB 1 AATTCGGACAGTATGAATACAGCGGATTAATACAGGTAAGTTTCATTACACGGAAGAAC 60
QY 61 CTTTAAAGAAATAATGAAGGTATTACCTTGTTGCCAACGGGAATGCTAAATATGCC 120
DB 61 CTTTAAAGAAATAATGAAGGTATTACCTTGTTGCCAACGGGAATGCTAAATATGCC 120
QY 121 GAGTTTTCTACTGAATACAGCAATCCAGCAATTCATTTCATATTTGACTGATGCTGAA 180
DB 121 GAGTTTTCTACTGAATACAGCAATCCAGCAATTCATTTCATATTTGACTGATGCTGAA 180
QY 181 TGTGGACTTTATAGATAATGACGATGAAGATTAAAT 217
DB 181 TGTGGACTTTATAGATAATGACGATGAAGATTAAAT 217
RESULT 2
A68928 217 bp DNA linear PAT 06-MAY-1999
LOCUS
DEFINITION
    Sequence 99 from Patent WO9802547.
ACCESSION
    A68928
VERSION
    A68928.1 GI:4759847
KEYWORDS
    unidentified
SOURCE
    unclassified.
ORGANISM
    unclassified.
REFERENCE
    1 (bases 1 to 217)
AUTHORS Nassif,X., Tinsley,C., Achtman,M., Ruelle,J., Vinals,C. and
    Merker,P.
TITLE
    DNA AND SPECIFIC PROTEINS OR PEPTIDES OF THE NEISSERIA MENINGITIDIS
    SPECIES BACTERIA, METHOD FOR OBTAINING THEM AND THEIR BIOLOGICAL
    APPLICATIONS
JOURNAL
    Patent: WO 9802547-A 99 22-JAN-1998;
    INST NAT SANTE RECH MED (FR)
COMMENT
    Other publication FR 2751000 19980116.
FEATURES
    source
        Location/Qualifiers
            1..217
                /organism="unidentified"
                /mol_type="genomic DNA"
                /db_xref="taxon:32644"
BASE COUNT 76 a 31 c 44 g 66 t
ORIGIN
    Query Match 100.0%; Score 217; DB 6; Length 217;
    Best Local Similarity 100.0%; Pred. No. 8.1e-44;
    Matches 217; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AATTCGGACAGTATGAATACAGCGGATTAATACAGGTAAGTTTCATTACACGGAAGAAC 60
DB 1 AATTCGGACAGTATGAATACAGCGGATTAATACAGGTAAGTTTCATTACACGGAAGAAC 60

```

```

QY 61 CTTTAAAGAAATAATGAAGGTATTACCTTGTTGCCAACGGGAATGCTAAATATGCC 120
DB 61 CTTTAAAGAAATAATGAAGGTATTACCTTGTTGCCAACGGGAATGCTAAATATGCC 120
QY 121 GAGTTTTCTACTGAATACAGCAATCCAGCAATTCATTTCATATTTGACTGATGCTGAA 180
DB 121 GAGTTTTCTACTGAATACAGCAATCCAGCAATTCATTTCATATTTGACTGATGCTGAA 180
QY 181 TGTGGACTTTATAGATAATGACGATGAAGATTAAAT 217
DB 181 TGTGGACTTTATAGATAATGACGATGAAGATTAAAT 217
RESULT 3
BD063028 217 bp DNA linear PAT 27-AUG-2002
LOCUS
DEFINITION
    DNA and specific proteins or peptides of the Neisseria meningitidis
    species bacteria, method for obtaining them and their biological
    applications.
ACCESSION
    BD063028
VERSION
    BD063028.1 GI:22608631
KEYWORDS
    JP 2001504684-A/90.
SOURCE
    unidentified
ORGANISM
    unclassified.
REFERENCE
    1 (bases 1 to 217)
AUTHORS Nassif,X., Tinsley,C., Achtman,M., Ruelle,J.L., Vinals,C. and
    Merker,P.
TITLE
    DNA and specific proteins or peptides of the Neisseria meningitidis
    species bacteria, method for obtaining them and their biological
    applications
JOURNAL
    Patent: JP 2001504684-A 90 10-APR-2001;
    INSTITUT NATIONAL DE LA SANTE ET DE LA RECHERCHE MEDICALE, MAX
    PLANCK GESELLSCHAFT ZUR FORDERUNG DER WISSENSCHAFTEN EV BERLIN,
    SMITHKLINE BEECHAM
COMMENT
    PN JP 2001504684-A/90
    PD 10-APR-2001
    PF 11-JUL-1997 JP 1998505685
    PR 12-JUL-1996 FR 96/08768
    PI XAVIER NASSIF, COLIN TINSLEY, MARK ACHTMAN, JEAN LOUIS RUELLE, PI
    CARLA VINALS.
    PI PETRA MERKER
    PC C12N15/31,C07K14/22,C07K16/12,A61K39/095,C12Q1/68,G01N33/53 CC
    Strandedness: Single;
    CC Topology: Linear;
    FH Key Location/Qualifiers.
FEATURES
    source
        1..217
            /organism="unidentified"
            /mol_type="genomic DNA"
            /db_xref="taxon:32644"
BASE COUNT 76 a 31 c 44 g 66 t
ORIGIN
    Query Match 100.0%; Score 217; DB 6; Length 217;
    Best Local Similarity 100.0%; Pred. No. 8.1e-44;
    Matches 217; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AATTCGGACAGTATGAATACAGCGGATTAATACAGGTAAGTTTCATTACACGGAAGAAC 60
DB 1 AATTCGGACAGTATGAATACAGCGGATTAATACAGGTAAGTTTCATTACACGGAAGAAC 60
QY 61 CTTTAAAGAAATAATGAAGGTATTACCTTGTTGCCAACGGGAATGCTAAATATGCC 120
DB 61 CTTTAAAGAAATAATGAAGGTATTACCTTGTTGCCAACGGGAATGCTAAATATGCC 120
QY 121 GAGTTTTCTACTGAATACAGCAATCCAGCAATTCATTTCATATTTGACTGATGCTGAA 180
DB 121 GAGTTTTCTACTGAATACAGCAATCCAGCAATTCATTTCATATTTGACTGATGCTGAA 180
QY 181 TGTGGACTTTATAGATAATGACGATGAAGATTAAAT 217
DB 181 TGTGGACTTTATAGATAATGACGATGAAGATTAAAT 217

```

```

RESULT 4
NMA222491
LOCUS
DEFINITION
Neisseria meningitidis serogroup A strain Z491 complete genome;
segment 2/7.
ACCESSION
AL162753 AL157959
VERSION
AL162753.2 GI:7379120
KEYWORDS
Neisseria meningitidis Z2491
SOURCE
Neisseria meningitidis Z2491
ORGANISM
Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
Neisseriaceae; Neisseria.
REFERENCE
1 (bases 1 to 349061)
Parkhill,J., Achtman,M., James,K.D., Bentley,S.D., Churcher,C.,
Klee,S.R., Morelli,G., Basham,D., Brown,D., Chillingworth,T.,
Davies,S.R., Davis,P., Devlin,K., Feltwell,T., Hamlin,N.,
Holroyd,S., Jagels,K., Leather,S., Moule,S., Mungall,K.,
Quail,M.A., Rajandream,M.A., Rutherford,K.M., Simmonds,M.,
Skellton,J., Whitehead,S., Spratt,B.G. and Barrall,B.G.
Complete DNA sequence of a serogroup A strain of Neisseria
meningitidis Z2491
Nature 404 (6777), 502-506 (2000)
JOURNAL
NATURE 404 (6777), 502-506 (2000)
MEDLINE
20222556
PUBMED
10761919
REFERENCE
2 (bases 1 to 349061)
Parkhill,J.
Direct Submission
Submitted (30-MAR-2000) Submitted on behalf of the Neisseria
sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk
AUTHORS
Notes:
Details of N. meningitidis sequencing at the Sanger Centre are
available on the Sanger wide web.
(URL, http://www.sanger.ac.uk/Projects/N_meningitidis/).
FEATURES
source
1..349061
/organism="Neisseria meningitidis Z2491"
/mol_type="Genomic DNA"
/strain="Z2491"
/db_xref="taxon:122587"
/note="serogroup: A"
complement(24..206)
/label="ATR repeat; hmms hit to HMM ATR (1 - 183), score:
310.39"
RBS
209..212
Gene
220..681
CDS
220..681
/gene="NMA0368"
/note="NMA0368, probable integral membrane protein, len:
153 aa; contains four probable transmembrane domains"
/codon_start=1
/transl_table=11
/product="putative integral membrane protein"
/protein_id="CAB83669.1"
/db_xref="GI:7379121"
/transl_table=11
/db_xref="SPTREMBL:Q9JH7"
/translation="MQQNRKSPFIYVNLVSVLWIASLSNVAFYLGNGHSGEGLTV
LILGTSASLDIRVCAYANFVMDAAIVLLALRKVYFVHAAPFGLVAFSVKAVYV
DEAGTSDIVRYGAGFYLWFAAFASIGTFAGKNKKAASADGTNDV"
671..1492
/gene="hemK"
671..1492
/gene="hemK"
/note="NMA0369, hemK, HemK protein, len: 273 aa; similar
to e.g. HEMK_ECOLI P37186 HEMK protein (277 aa), fasta
scores; E(): 0, 42.3% identity in 279 aa overlap. Contains
PS00092 N-6 Adenine-specific DNA methylases signature"
/codon_start=1
/transl_table=11
/gene="hemK"
/note="Core DNA uptake sequence: gccgtctgaa"
1162..1171
/gene="hemK"
/label="DUS"
1187..1207
/gene="hemK"
/note="PS00032 N-6 Adenine-specific DNA methylases
signature"
/note="Core DNA uptake sequence: gccgtctgaa"
1565..1568
/label="DUS"
1578..2966
/gene="NMA0370"
1578..2966
/gene="NMA0370"
/note="NMA0370, probable integral membrane protein, len:
462 aa; similar to hypothetical proteins e.g. Y325_HAEIN
P44640 hypothetical protein H10325 (450 aa), fasta
scores; E(): 0, 49.8% identity in 464 aa overlap"
/codon_start=1
/transl_table=11
/product="putative integral membrane protein"
/protein_id="CAB83671.1"
/db_xref="GI:7379123"
/db_xref="SPTREMBL:Q9JH5"
/translation="MNAVVAIVVIMLVLSRVHVVLSTIGAFVGAVAGMPLQNI
DAAGQVSQAGIIPFNKGLGEGKIALSYAMLGAFAMAITHSGLPQOLAGAVVRKLN
RGMPDSVRGEGVYVWLLSIIIVMGMSQNIPIPIAFIMVIVPLLVNRLKTR
GLIACVITFTGLVTVTFMFLPYGCAIFLNLGNTHSAAPQOLDVKNINWMAAIPAL
GMLAGLLAFVHYKRPRLYOSNNADTGNADANRPPQSAYSRLAAVAIVCFALQL
WEDSVILGAMLGFAVFMGLVINRDKANDVFEGIKMAMVGFIMAAQGFVAVYVGL
TGIQIPLVSSMAIPGNSKMAALVIVVGLLVTMIGSGSFSTLPIIAIYVPLCVGL
GFSPLATVAIVGTAGAGDAGSPASDSTLGTMLGNADQGHDIRDSVPTFIHYNIP
LTIAGWIAAVL"
2967..3191
/gene="slyX"
2967..3191
/gene="slyX"
/note="NMA0371, slyX, SLX protein homolog, len: 74 aa;
similar to SLX_HAEIN P44759 SLX protein homolog (73 aa),
fasta scores; E(): 0.0017, 33.8% identity in 74 aa
overlap, and SLX_ECOLI P30857 SLX protein (72 aa), fasta
scores; E(): 0.24, 32.4% identity in 68 aa overlap"
/codon_start=1
/transl_table=11
/product="SLX protein homolog"
/protein_id="CAB83672.1"
/db_xref="GI:7379124"
/db_xref="SPTREMBL:Q9JH4"
/translation="NDAYCEPHRITETLEIQSALQEDYIAGLNAMVAELRQTLIDLQQA
QRLLYQKQQRNPDAQPQPSLRDEIPPHY"
complement(3271..3423)
/gene="NMA0372"
complement(3271..3423)
/gene="NMA0372"
/note="NMA0372, unknown, questionable CDS, len: 50 aa"
/codon_start=1
/transl_table=11
/product="very hypothetical protein NMA0372"

```

misc\_feature  
/protein\_id="CAB83673.1"  
/db\_xref="GI:7379125"  
/db\_xref="SPTREMBL:Q9JWH3"  
/translation="MFCRLSPGSDGFLAVYVPVGFICLQKGTAYFRSGGKRPHRI  
KGILSD"  
3562..3571  
/note="Core DNA uptake sequence: gccgtctgaa"  
/label=DUS  
complement(3586..4356)  
/gene="thiP"  
complement(3586..4356)  
/gene="thiP"  
/note="NMA0373, thiP, probable ThiF protein, len: 256 aa;  
similar to e.g. THIF\_ECOLI P30138 ThiF protein (251 aa),  
fasta scores; E(): 0, 43.1% identity in 246 aa overlap,  
and MOEB\_ECOLI P12282 molybdopterin biosynthesis MOEB  
protein\_2 (249 aa), fasta scores; E(): 0, 43.9% identity in  
244 aa overlap (note that N.m. does not have orthologs of  
any other molybdopterin biosynthesis proteins). Contains  
pfam match to entry PF00899 ThiF\_family, ThiF family"  
/codon\_start=1  
/transl\_table=1  
/product="ThiF protein"  
/protein\_id="CAB83674.1"  
/db\_xref="GI:7379126"  
/db\_xref="SPTREMBL:Q9JWH2"  
/translation="MTTEHDNDFAFLRYSRHLLDEIGIEGQOKLSAAHILVSGG  
GLGAALFYLAASGIGTITADSDTVLHNLQVAFDEGDKLKEALADRHHN  
HTVDVRIINELKIDGRLTVQADIVLDCDNYATQAVNRACVQAKTEPLVSGAAVR  
FEGOLAVRPLDPSVACVLFDCGSGDGLFSGVFSPLVGIIGSTQAAEALKILL  
DAGPSHGRILAVYRALEGGMQYFPLRNPCFVCGAER"  
3787..3796  
/note="Core DNA uptake sequence: gccgtctgaa"  
/label=DUS  
complement(3847..4257)  
/gene="thiP"  
/note="Pfam match to entry PF00899 ThiF\_family, ThiF  
family, score 186.60, E-value 4.1e-52"  
complement(4418..4427)  
/note="Core DNA uptake sequence: gccgtctgaa"  
/label=DUS  
4471..7224  
/gene="ppc"  
4471..7224  
/gene="ppc"  
/EC\_number="4.1.1.31"  
/note="NMA0374, ppc, phosphoenolpyruvate carboxylase, len:  
917 aa; similar to many e.g. CAPP\_RHOFA Q32483  
phosphoenolpyruvate carboxylase (EC 4.1.1.31) (936 aa),  
fasta scores; E(): 0, 43.3% identity in 928 aa overlap.  
Contains 2x Pfam match to entry PF00311 PEPcase,  
Phosphoenolpyruvate carboxylase, PS00017 ATP/GTP-binding  
site motif A (P-loop), and PS00393 Phosphoenolpyruvate  
carboxylase active site 2"  
/codon\_start=1  
/transl\_table=1  
/product="phosphoenolpyruvate carboxylase"  
/protein\_id="CAB83675.1"  
Query Match 88.2%; Score 191.4; DB 1; Length 349061;  
Best Local Similarity 98.2%; Pred. No. 5.5e-38;  
Matches 215; Conservative 0; Mismatches 1; Indels 3; Gaps 2;  
QY 1 AATTCGACAGTATGAATACAGCGGATTAATACAGGTAAGTTTCATTACACGGAAGAAC 60  
DB 59922 AATTCGACAGTATGAATACAGCGGATTAATACAGGTAAGTTTCATTACACGGAAGAAC 59981  
QY 61 CTTTAAAGAAATATGAAGTATACCTGTTTCCCAACGG--GAATGTAATATGTC 118  
DB 59982 CTTTAAAGAAATATGAAGTATACCTGTTTCCCAACGGCGAAGTGTGAATATGTC 60041  
QY 119 CCGAGTTTTCTACTGAATACGGAATCCAGCCATTTCATTATTCAGTGGATGGCTG 178

Db 60042 CCGAGTTTTCTACTG-ATAGCGAATCCAGCCATTTCATTATTCAGTGGATGGCTG 60100  
QY 179 AATTCGACAGTATGAATACAGCGGATTAATACAGGTAAGTTTCATTATTCAGTGGATGGCTG 217  
DB 60101 AATTCGACAGTATGAATACAGCGGATTAATACAGGTAAGTTTCATTATTCAGTGGATGGCTG 60139  
RESULT 5  
AX024058 858 bp DNA linear PAT 15-SEP-2000  
LOCUS  
DEFINITION Sequence 1 from Patent FR2785293.  
ACCESSION AX024058  
VERSION AX024058.1 GI:10184370  
KEYWORDS  
SOURCE  
ORGANISM  
Neisseria meningitidis  
Neisseria meningitidis  
Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
Neisseriaceae; Neisseria.  
REFERENCE  
1  
AUTHORS Nassif X., Tinsley C., Aujame L., Perrin A., Rokbi B.,  
Bouchardon A. and Renauld M.G.  
JOURNAL Patent: FR 2785293-A 1 05-MAY-2000;  
PASTEUR MERIEUX SERUMS VACC (FR)  
FEATURES  
source  
1..858  
/organism="Neisseria meningitidis"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:487"  
1..858  
/note="unnamed protein product"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="CAC08905.1"  
/db\_xref="GI:10184371"  
/translation="MSEKLMSPETVIEHLGVKWSHTVPAALAIANAYDACATE  
VEVRLDPDEKIVIKDNGICMSDEINDVLRIGRNEREKOASPGRIPTGKGLG  
KUALFGLNKEIESTIQGNERNVIFLDVAEIRRSKGIQPERKESVESNESCTTIT  
LTELTKQGYPLDNYVHLSRLFPDPAQDKIKVSLNGSEPKLIDGNLKYLDVTPQFE  
WEYQDLATNIISSLSKFPQYEGSLIQGKFTTEKPLKNMKVIGITLFFANGRWNVPEF  
FTDSBSHF"  
BASE COUNT 297 a 145 c 182 g 234 t  
ORIGIN  
Query Match 60.1%; Score 130.4; DB 6; Length 858;  
Best Local Similarity 97.5%; Pred. No. 2.4e-22;  
Matches 154; Conservative 0; Mismatches 1; Indels 3; Gaps 2;  
QY 1 AATTCGACAGTATGAATACAGCGGATTAATACAGGTAAGTTTCATTACACGGAAGAAC 60  
DB 701 AATTCGACAGTATGAATACAGCGGATTAATACAGGTAAGTTTCATTACACGGAAGAAC 760  
QY 61 CTTTAAAGAAATATGAAGTATACCTGTTTCCCAACGG--GAATGTAATATGTC 118  
DB 761 CTTTAAAGAAATATGAAGTATACCTGTTTCCCAACGGCGAAGTGTGAATATGTC 820  
QY 119 CCGAGTTTTCTACTGATAGCGAATCCAGCCATTTCCTA 156  
DB 821 CCGAGTTTTCTACTG-ATAGCGAATCCAGCCATTTCCTA 857  
RESULT 6  
AX024111 858 bp DNA linear PAT 15-SEP-2000  
LOCUS  
DEFINITION Sequence 54 from Patent FR2785293.  
ACCESSION AX024111  
VERSION AX024111.1 GI:10184422  
KEYWORDS  
SOURCE  
ORGANISM  
Neisseria gonorrhoeae  
Neisseria gonorrhoeae  
Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
Neisseriaceae; Neisseria.  
REFERENCE  
1  
AUTHORS Nassif X., Tinsley C., Aujame L., Perrin A., Rokbi B.,



Bouchardon, A. and Renauld, M.G.  
 Patent: FR 2785293-A 54 05-MAY-2000;  
 PASTEUR MERIEUX SERUMS VACC (FR)  
 Location/Qualifiers  
 source 1..858  
 /organism="Neisseria gonorrhoeae"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:485"  
 CDS 1..858  
 /note="unnamed protein product"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAC08931.1"  
 /db\_xref="GI:10184423"  
 /translation="MSEELKMSFPTVIEHLGVKMYSHVTPAIAELIANAYDACATE  
 VEVRLFDKEHIVIKDNGIGMSFDEINDVLRIGNRREKQASFGCRIPITGKGLG  
 KIALPLGNKIKISITOGNRYVTFILDVAIEKISRIQPERKQESRPNTENGTTIT  
 LTELPLKQYPLDINVTGHSRLFDPAQDFKIKVLSNGSEPRIDNKLKYNLVTPOFE  
 WYQDIAINISLSKFEQREYSGIQQGFIITTEKPLKNNMKGITLFGANGRVNMPF  
 FTDSSSHF"  
 BASE COUNT 302 a 151 c 178 g 227 t  
 ORIGIN  
 Query Match 60.1%; Score 130.4; DB 6; Length 858;  
 Best Local Similarity 97.5%; Pred. No. 2.4e-22;  
 Matches 154; Conservative 0; Mismatches 1; Indels 3; Gaps 2;  
 QY 1 AATCGACAGTATGAATACAGCGATTAAATCAAGTAAAGTTCATTACACGGGAAAAAC 60  
 |||||  
 Db 701 AATTCGAACAGTATGAATACACGGATTATATCAAGTAAAGTTCATTACACGGGAAAAAC 760  
 |||||  
 QY 61 CTTTAAAGTATATGAAGTATATACCTTTGTCGCAACGG--GAATGTAATATATGC 118  
 |||||  
 Db 761 CTTTAAAGTATATGAAGTATATACCTTTGTCGCAACGGGAAATGTAATATATGC 820  
 |||||  
 QY 119 CCGAGTTTTCCTGCTAATGAGCAATCCAGCCATTCTTA 156  
 |||||  
 Db 821 CCGAGTTTTCCTGCTAATGAGCAATCCAGCCATTCTTA 857  
 |||||  
 RESULT 7  
 AC106472/c  
 LOCUS  
 DEFINITION Rattus norvegicus clone CH230-15201, \*\*\*.SEQUENCING IN PROGRESS  
 \*\*\* 3 unordered pieces.  
 AC106472  
 ACCESSION  
 VERSION HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.  
 KEYWORDS Rattus norvegicus (Norway rat)  
 SOURCE  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 1 (bases 1 to 223954)  
 Muzny, D., Marie, Metzker, M., Lee, Abramson, S., Adams, C., Alder, J.,  
 Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,  
 Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,  
 Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,  
 Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,  
 Bryant, N., Bunay, C., Burch, P., Burrell, K., Calderon, E.,  
 Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,  
 Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,  
 Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,  
 Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,  
 Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,  
 Draper, H., Dugan-Rocha, S., Dunn, A., Dutkin, K., Duval, B., Evans, K.,  
 Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,  
 Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,  
 Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,  
 Gebregiorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,  
 Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,  
 Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,  
 Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M.,

Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,  
 Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,  
 Karpathy, S., Kelly, S., Khan, Z., King, L., Kovar, C.,  
 Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,  
 Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,  
 Lorensuhewa, L., Loulsecq, H., Lozado, R.J., Lu, X., Ma, J.,  
 Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A.,  
 Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,  
 Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,  
 Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,  
 Morgan, M., Morris, K., Morris, S., Muridasa, M., Murphy, M., Nair, L.,  
 Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,  
 Nwaekeleneh, O., Okwuonu, G., Olarunpunsagoon, A., Pal, S., Parks, K.,  
 Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,  
 Plopper, F., Polidexter, A., Popovic, D., Primus, E., Pu, L.L.,  
 Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,  
 Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,  
 Rives, C., Rokey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,  
 Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,  
 Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D.,  
 Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,  
 Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,  
 Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,  
 Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,  
 Wang, C., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,  
 Williams, G., Willson, R., Wleczyk, R., Woodden, H., Worley, K.,  
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,  
 Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von  
 Niederhauser, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,  
 Weinstein, G. and Gibbs, R.A.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 223954)  
 Worley, K.C  
 Direct Submission  
 Submitted (12-JAN-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 223954)  
 Rat Genome Sequencing Consortium.  
 Direct Submission  
 Submitted (13-MAY-2003) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On May 13, 2003 this sequence version replaced gi:23265983.  
 The sequence in this assembly is a combination of BAC-based reads  
 and whole genome shotgun sequencing reads assembled using Atlas  
 (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described  
 in the feature table below represents a scaffold in the Atlas  
 assembly (a 'contig-scaffold'). Within each contig-scaffold,  
 individual sequence contigs are ordered and oriented, and separated  
 by sized gaps filled with Ns to the estimated size. The sequence  
 may extend beyond the ends of the clone and there may be sequence  
 contigs within a contig-scaffold that consist entirely of whole  
 genome shotgun sequence reads. Both end sequences and whole genome  
 shotgun sequence only contigs will be indicated in the feature  
 table.

----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
 ----- Project Information  
 Center project name: GKX  
 Center clone name: CH230-15201  
 ----- Summary Statistics  
 Assembly program: Atlas 3.0;  
 Consensus quality: 20381 bases at least Q40  
 Consensus quality: 20755 bases at least Q30  
 Consensus quality: 210432 bases at least Q20  
 Estimated insert size: 215302; sum-of-contigs estimation  
 Quality coverage: 6x in Q20 bases; sum-of-contigs estimation  
 -----

\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 3 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 221228: contig of 221228 bp in length  
2 221329: gap of unknown length  
3 221329: contig of 1096 bp in length  
4 222426: gap of unknown length  
5 222525: contig of 1430 bp in length.

Location/Qualifiers  
1. 223954  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-15201"  
/note="clone boundary"  
clone\_end:596  
site:EcoRI  
end\_sequence:BH335992"  
1514..1860  
/note="clone boundary"  
clone\_end:17  
site:EcoRI  
end\_sequence:BH335441"  
95383..97199  
/note="wgs\_end\_extension"  
clone\_end:17"  
BASE COUNT 62554 a 48124 c 45176 g 56913 t 11187 others  
ORIGIN

Query Match 19.7%; Score 42.8; DB 2; Length 223954;  
Best Local Similarity 53.6%; Pred. No. 0.63;  
Matches 89; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

42 TTCATTACACGAAACCTTTAAGATATATGAAGGTATACCTCTTTGCCAAC 101  
131056 TTCATTGAGAGCATGCTATATAACAAAGATTCAAATTTATCATATTCACAT 130997

102 GGAATGTTAATATGCCGAGTTTTCACCTGAATAGCAATCCAGCATTTCTATTCAT 161  
130996 GAAATAGACAACTGCCCACTTTTAAATAATAAATAGTCCATTTATTTTGT 130937

162 ATTGACTGGATGGCTGAATGTGGACTTTATAGATATGACATGA 207  
130936 TTCATTTTAAATTTAGTTTGGAGATCCATAGCTAATGATGATGA 130891

RESULT 8  
AC133419/c  
LOCUS  
DEFINITION Rattus norvegicus clone CH230-52418, \*\*\* SEQUENCING IN PROGRESS  
\*\*\*, 4 unordered pieces.  
ACCESSION AC133419  
VERSION AC133419.2 GI:25007534  
KEYWORDS HTG, HTGS PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.  
SOURCE Rattus norvegicus (Norway rat)  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
1 (bases 1 to 261007)  
Muzny, D. Marie, Metzker, M. Lee, Abramson, S., Adams, C., Alder, J.,  
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,  
Anyalebechi, V., Ayodeji, A., Ayodeji, M., Baca, E., Baden, H.,  
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,  
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,

Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,  
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,  
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,  
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,  
Devilla, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,  
Deigado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,  
Egan, A., Escotto, M., Eugene, C., Evans, A., Falls, T., Fan, G.,  
Fraser, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,  
Freaner, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,  
Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,  
Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,  
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,  
Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M.,  
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idiebird, P., Jackson, A.,  
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,  
Karpathy, S., Kelly, S., Khan, Z., King, L., Kovar, C.,  
Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,  
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,  
Lorensuhewa, L., Loulseghe, H., Lozano, R.J., Lu, X., Ma, J.,  
Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,  
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,  
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,  
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,  
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,  
Nwaokeme, O., Okwuonu, G., Olarunpungu, A., Pal, S., Parks, K.,  
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,  
Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L.,  
Pua, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,  
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,  
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,  
Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,  
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajls, D.,  
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,  
Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,  
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,  
Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,  
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,  
Williams, G., Willson, R., Wleciyk, R., Wooden, H., Worley, K.,  
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,  
Yu, F., Zhang, J., Zhou, J., Zhou, S., Zhao, S., Dunn, D., von  
Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,  
Weinstock, G. and Gibbs, R.A.

Direct Submission  
Unpublished  
2 (bases 1 to 261007)  
Rat Genome Sequencing Consortium.  
Submitted (12-SEP-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 261007)  
Rat Genome Sequencing Consortium.  
Direct Submission  
Submitted (15-NOV-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
The sequence in this assembly is a combination of BAC based reads  
and whole genome shotgun sequencing reads assembled using Atlas  
(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described  
in the feature table below represents a scaffold in the Atlas  
assembly (a 'contig-scaffold'). Within each contig-scaffold,  
individual sequence contigs are ordered and oriented, and separated  
by sized gaps filled with Ns to the estimated size. The sequence  
may extend beyond the ends of the clone and there may be sequence  
contigs within a contig-scaffold that consist entirely of whole  
genome shotgun sequence reads. Both end sequences and whole genome  
shotgun sequence only contigs will be indicated in the feature  
table.

----- Genome Center  
Center: Baylor College of Medicine

Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information -----  
 Center project name: KAYS  
 Center clone name: CH230-52418  
 ----- Summary Statistics -----  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 223700 bases at least Q40  
 Consensus quality: 226843 bases at least Q30  
 Consensus quality: 228976 bases at least Q20  
 Estimated insert size: 228564; sum-of-contigs estimation  
 Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

\*\*\*\*\* NOTE: Estimated insert size may differ from sequence length  
 (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 4 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

\*\*\*\*\*  
 1 240201: contig of 240201 bp in length  
 \* 240202 240301: gap of unknown length  
 \* 240302 256939: contig of 16638 bp in length  
 \* 256940 257039: gap of unknown length  
 \* 257040 258335: contig of 1496 bp in length  
 \* 258336 258635: gap of unknown length  
 \* 258636 261007: contig of 2372 bp in length.

FEATURES  
 source  
 1..261007  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10116"  
 /clone="CH230-52418"  
 1..1542  
 /note="wgs contig"  
 misc\_feature  
 1593..4461  
 /note="wgs contig"  
 misc\_feature  
 252331..253535  
 /note="wgs contig"  
 misc\_feature  
 255445..256939  
 /note="wgs contig"

BASE COUNT 68651 a 52580 c 49193 g 59942 t 30641 others  
 ORIGIN

Query Match 19.7%; Score 42.8; DB 2; Length 261007;  
 Best Local Similarity 53.6%; Pred. No. 0.61;  
 Matches 89; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 42 TTCAATACACGGAAACCTTTAAAGAAATATATGAAGGATATACCTTGTGGCCAAAC 101  
 Db TCAATTCGGAGGACGATGCTATATAACAAAGATCAAAATTTACTATCATATTTCCACAT 95210

QY 102 GGAATGTAATAATGCCGAGTTTTCACATGAATAGCAATCCAGCCATTTCTATTCAT 161  
 Db GAAATAGACAAACTGCCCACTTTTAAATAATAAATAGTCCATTTATTTTGT 95150

QY 162 ATTGACCTGGATGCTGATGCTATATAGATATAGCATGA 207  
 Db TCAATTTTAAATTTAGTTTGGAGAAATCCCATAGCTAATGATGATGA 95104

RESULT 9  
 AC129204  
 LOCUS AC129204 163284 bp DNA linear ROD 23-FEB-2003  
 DEFINITION Mus musculus chromosome 12 clone RP24-230P5, complete sequence.  
 ACCESSION AC129204  
 VERSION AC129204.4 Gi:28475615  
 KEYWORDS HTG.  
 SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 163284)  
 McPherson,J.D. and Waterston,R.H.  
 TITLE The sequence of Mus musculus clone  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 163284)  
 McPherson,J.D. and Waterston,R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (27-JUL-2002) Genome Sequencing Center, 4444 Forest Park  
 Parkway, St. Louis, MO 63108, USA  
 3 (bases 1 to 163284)  
 McPherson,J.D. and Waterston,R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (13-FEB-2003) Genome Sequencing Center, 4444 Forest Park  
 Parkway, St. Louis, MO 63108, USA  
 4 (bases 1 to 163284)  
 McPherson,J.D. and Waterston,R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (23-FEB-2003) Genome Sequencing Center, 4444 Forest Park  
 Parkway, St. Louis, MO 63108, USA  
 COMMENT On Feb 23, 2003 this sequence version replaced gi:28372744.

FEATURES  
 Location/Qualifiers  
 1..163284  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 /chromosome="12"  
 /clone="RP24-230P5"  
 BASE COUNT 52814 a 29602 c 29566 g 51302 t  
 ORIGIN

Query Match 19.1%; Score 41.4; DB 10; Length 163284;  
 Best Local Similarity 56.1%; Pred. No. 1.5;  
 Matches 78; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 27 TTAATCAAGGTAAGTTCATTACACGGAACCTTTAAAGAAATATGAAAGGATTT 86  
 Db TTAATAATAGTCACATTCATTAAAGCTTAGTCACCTTGATATTAACACCATATCTT 148626

QY 87 ACCTCTTTGCCAACGGGAATGGTAAATATGCCGAGTTTTCCTGATAGCATCCA 146  
 Db TCCATATTTGTTTAGTGTATCTCAATATATATAGGTTTCAATGAAACAAGTATCCA 148696

QY 147 GCATTTCTATTCAATTTT 165  
 Db TACAATTAATAACAATTT 148705

RESULT 10  
 AC125488  
 LOCUS AC125488 183202 bp DNA linear ROD 21-DEC-2002  
 DEFINITION Mus musculus chromosome 12 clone RP23-75M24, complete sequence.  
 ACCESSION AC125488  
 VERSION AC125488.3 Gi:27356773  
 KEYWORDS HTG.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 183202)  
 McPherson,J.D. and Waterston,R.H.  
 TITLE The sequence of Mus musculus clone

Unpublished  
2 (bases 1 to 183202)  
REFERENCE  
McPherson,J.D. and Waterston,R.H.  
TITLE  
Direct Submission  
JOURNAL  
Submitted (27-JUN-2002) Genome Sequencing Center, 4444 Forest Park  
Parkway, St. Louis, MO 63108, USA  
3 (bases 1 to 183202)  
REFERENCE  
McPherson,J.D. and Waterston,R.H.  
TITLE  
Direct Submission  
JOURNAL  
Submitted (06-DEC-2002) Genome Sequencing Center, 4444 Forest Park  
Parkway, St. Louis, MO 63108, USA  
4 (bases 1 to 183202)  
REFERENCE  
McPherson,J.D. and Waterston,R.H.  
TITLE  
Direct Submission  
JOURNAL  
Submitted (21-DEC-2002) Genome Sequencing Center, 4444 Forest Park  
Parkway, St. Louis, MO 63108, USA  
On Dec 21, 2002 this sequence version replaced gi:26080599.  
COMMENT  
----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: http://genome.wustl.edu/gsc/index.shtml  
Contact: submissions@wustl.edu  
----- Project Information -----  
Center project name: M\_BA0075M24  
----- Location/Qualifiers -----  
1. 183202  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="12"  
/clone="RP23-75M24"  
BASE COUNT 60541 a 33423 c 33788 g 55450 t  
ORIGIN  
Query Match 19.1%; Score 41.4; DB 10; Length 183202;  
Best Local Similarity 56.1%; Pred. No. 1.4;  
Matches 78; Conservative 0; Mismatches 61; Indels 0; Gaps 0;  
QY 27 TTATACAGGTAAGTTCATTACACGGAAGAACTTTAAGATATATATGAAGGTATT 86  
Db 12250 TTAATAATAGTGACATTCATTAAAGCTTAGTACCCCTTGGATATTAACACATATCTT 12309  
QY 87 ACCTTGTTTGGCAACCGGAATGGTAAATATGCGGAGTTTTCAGTGAATAGCGAATCCA 146  
Db 12310 TCCATATTGTTAGTGTATACATCATATATATAGTTTACATGAAACAGTATTC 12369  
QY 147 GCCATTTCATCATATT 165  
Db 12370 TACAATTATTATACATTT 12388  
RESULT 11  
AC129541/1  
LOCUS 198303 bp DNA linear HTG 22-MAY-2003  
DEFINITION Mus musculus clone RP23-248M12, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 4  
unordered pieces.  
ACCESSION AC129541  
VERSION AC129541.4 GI:30984790  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_FULLTOP; HTGS\_ACTIVEFIN.  
SOURCE  
Mus musculus (house mouse)  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 198303)  
REFERENCE  
Birken,B., Nusbaum,C. and Lander,E.  
TITLE  
Unpublished  
JOURNAL  
Submitted (21-DEC-2002) Genome Sequencing Center, 4444 Forest Park  
Parkway, St. Louis, MO 63108, USA  
2 (bases 1 to 198303)  
REFERENCE  
Birken,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,  
Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,  
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collamore,A.,  
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Gage,D., Galagan,J.,  
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,  
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kanat,A.,  
Karatas,A., Kells,C., Lander,E., Lander,R., Lindblad-Toh,K.,  
Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,  
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mieng,V.,  
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,  
O'Connor,R., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,  
Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rogov,P.,  
Roman,J., Schauer,S., Schuback,R., Seaman,S., Severy,P., Smith,C.,  
Spencer,C., Stange-Thomann,N., Stojanovic,N., Talamas,J.,  
Tefaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,  
Zembek,L., Zimmer,A. and Zody,M.  
Direct Submission  
Submitted (30-JUL-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 198303)  
REFERENCE  
Birken,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,  
Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,  
Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,  
Collamore,A., Cook,A., Cooke,P., Corum,B., Dearellano,K.,  
Diaz,J.S., Dodge,S., Dooley,K., Dorris,B., Erickson,J., Faro,S.,  
Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S.,  
Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,  
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,  
Kamat,A., Karatas,A., Kells,C., Lander,E., Levine,R.,  
Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,  
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,  
Meldrim,J., Meneus,L., Mihova,T., Mieng,V., Murphy,T., Naylor,J.,  
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,  
O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,  
Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rogov,P.,  
Roman,J., Schauer,S., Schuback,R., Seaman,S., Severy,P., Smith,C.,  
Spencer,C., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,  
Talamas,J., Tefaye,S., Theodore,J., Topham,K., Travers,M.,  
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,  
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
Direct Submission  
Submitted (22-MAY-2003) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On May 22, 2003 this sequence version replaced gi:29126476.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center -----  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information -----  
Center project name: L26920  
Center clone name: 248\_M\_12  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 4 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence.  
\* as soon as it is available and the accession number will  
\* be preserved  
\* 1 82226: contig of 82226 bp in length  
\* 82227: gap of 100 bp  
\* 112802: contig of 30476 bp in length  
\* 112803: gap of 100 bp  
\* 112903: contig of 75221 bp in length  
\* 188124: gap of 100 bp  
\* 188224: contig of 10080 bp in length.  
Location/Qualifiers  
1. 198303  
/organism="Mus musculus"

```

/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="RP23-248M12"
/clone_lib="RPC1-23 Female Mouse BAC"

BASE COUNT      63714 a 36811 c 36425 g 60546 t      807 others
ORIGIN

Query Match      18.9%; Score 41; DB 2; Length 198303;
Best Local Similarity 49.5%; Pred. No. 1.8;
Matches 104; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

QY      6 GCACAGTATCAATACAGCGGATTATACAGGTAAAGTTCATTACACGGAAGAAACCTTTA 65
Db      155858 GCGAAGAAATATAAACCTCAGGTCGTAATATAAAAGACAGGAAAAACAAAACA 155799

QY      66 AGAATAATATGAAGAGTATACCTTGTGGTTCACGCGGAATGGTAATATGCCCGAGTT 125
Db      155798 AAACAAAAAANAANCAAAATAATACAAAGAAATAAATGAAATAGCTGGGTT 155739

QY      126 TTTCAGTGAATAGCGAATCCAGCCATTCTTATTCATATTTGACTGGATGGCTGATGTGG 185
Db      155738 TTTCAGAGATGAATAGATTGAAAATCTTATCCAAATTTACAAATTTGCGAGAAGAGA 155679

QY      186 ACTTTATAGATATGACGATGAAGATTTAA 215
Db      155678 AGATCCAAATTAATGAATTCAGATGAAA 155649

RESULT 12
AC131594
LOCUS      181815 bp      DNA      linear      HTG 26-FEB-2003
DEFINITION Mus musculus clone RP24-568P18, WORKING DRAFT SEQUENCE, 4 unordered
pieces.
ACCESSION AC131594
VERSION   AC131594.5 GI:28565729
KEYWORDS  HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE    Mus musculus (house mouse)
ORGANISM  Mus musculus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 181815)
AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
TITLE     Mus musculus, clone RP24-568P18
JOURNAL   Unpublished
REFERENCE 2 (bases 1 to 181815)
AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
           Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhalter,B.,
           Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
           Cook,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S.,
           Fero,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
           Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
           Hoxton,L., Hulme,W., Iilev,I., Johnson,R., Jones,C., Kamat,A.,
           Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
           Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
           McCarthy,M., Meldrim,J., Meneus,L., Minova,T., Mienga,V.,
           Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
           O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
           Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
           Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
           Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
           Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
           Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
           Zembek,L., Zimmer,A. and Zody,M.
DIRECT SUBMISSION
Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 181815)
AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
           Anderson,S., Archchi,H.M., Barna,N., Bastien,V., Bloom,T.,
           Boguslavskiy,L., Boukhalter,B., Camarata,J., Chang,J., Choepel,Y.,
           Collymore,A., Cook,A., Cooke,P., Corum,B., DeArelano,K.,
           Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Fero,S.,
           Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S.,

```

```

Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
Hall,J., Hoxton,L., Hulme,W., Iilev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., Maclean,C.,
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
Meldrim,J., Meneus,L., Minova,T., Mienga,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
DIRECT SUBMISSION
Submitted (26-FEB-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Feb 26, 2003 this sequence version replaced gi:28412028.
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green,P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seg.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L26731
Center clone name: 568_P_18
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 181080 bases at least Q40
Consensus quality: 181410 bases at least Q30
Consensus quality: 181465 bases at least Q20
Insert size: 184000; agarose-fp
Insert size: 181515; sum-of-contigs
Quality coverage: 9.0 in Q20 bases; agarose-fp
Quality coverage: 9.2 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 35214: contig of 35214 bp in length
* 35215 35314: gap of 100 bp
* 35315 60599: contig of 25285 bp in length
* 60600 60699: gap of 100 bp
* 60700 97829: contig of 37130 bp in length
* 97830 97929: gap of 100 bp
* 97930 181815: contig of 83886 bp in length.
Location/Qualifiers
1. 181815
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="RP24-568P18"
/clone_lib="RPC1-24 Male Mouse BAC"
misc_feature 1. 35214
/notes="assembly_fragment
clone_end:56
vector_side:left"
misc_feature 35315..60599
/notes="assembly_fragment"
misc_feature 60700..97829
/notes="assembly_fragment"
misc_feature 97930..181815
/notes="assembly_fragment"
FEATURES
source
misc_feature 1. 35214
/notes="assembly_fragment
clone_end:56
vector_side:left"
misc_feature 35315..60599
/notes="assembly_fragment"
misc_feature 60700..97829
/notes="assembly_fragment"
misc_feature 97930..181815
/notes="assembly_fragment"

```

```
clone_end:T7
vector_side:right
BASE COUNT 58499 a 33923 c 33630 g 55462 t 301 others
ORIGIN
Query Match 18.6%; Score 40.4; DB 2; Length 181815;
Best Local Similarity 49.5%; Pred. No. 2.6;
Matches 104; Conservative 0; Mismatches 106; Indels 0; Gaps 0;
QY 6 GGACAGTATGAATACAGCGGATTAAATACAGCTAAGTTCATTACAAACGGAACCTTTA 65
DB 173210 GGAAGAAATATAACCTAGGCTGAAATATAAAGACGGAACCAACAAACAA 173269
QY 66 AAGATAATATGAAGGATTACCTTGTTCCAAACGGGAATGTAATATGCCGAGTT 125
DB 173270 AAACAAAAAATAAAGAAACCAAAATAATACAAAGAAATAATGAACAAATAGCTGGTT 173329
QY 126 TTTCAGTGAATAGCAATCCAGCCATTCTATTCTATTTGACATGGATGGCTGAATCTGG 185
DB 173330 TTTCAGAGTGAATAGCAATCCAGCCATTCTATTCTATTTGACATGGATGGCTGAATCTGG 173389
QY 186 ACTTTATAGATAATGACGATGAAGATTAA 215
DB 173390 AGATCCAAATTAATGAATTAAGATGAAA 173419
RESULT 13
AP003188 301450 bp DNA linear BCT 10-JUL-2002
LOCUS Clostridium perfringens str. 13 DNA, complete genome, section 4/10.
DEFINITION AP003188 BA000016
ACCESSION AP003188.2 GI:18146726
VERSION
KEYWORDS
SOURCE Clostridium perfringens str. 13
ORGANISM Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
Clostridium.
REFERENCE
1 Shimizu,T., Ohtani,K., Hirakawa,H., Ohshima,K., Yamaehita,A.,
Shiba,T., Ogasawara,N., Hattori,M., Kuhara,S. and Hayashi,H.,
Complete genome sequence of Clostridium perfringens, an anaerobic
flesh-eater
Proc. Natl. Acad. Sci. U.S.A. 99 (2), 996-1001 (2002)
JOURNAL 21664373
MEDLINE 11792842
PUBMED
REFERENCE 2 (bases 1 to 301450)
Shimizu,T.
Direct Submission
Submitted (15-FEB-2001) Tohru Shimizu, Institute of Basic Medical
Sciences, University of Tsukuba, Department of Microbiology; 1-1-1
Tennohda, Tsukuba, Ibaraki 305-8575, Japan
(E-mail: tshimizu@md.tsukuba.ac.jp, Tel:81-298-53-3354,
Fax:81-298-53-3354)
COMMENT On Jan 14, 2002 this sequence version replaced gi:18144387.
FEATURES
Location/Qualifiers
1..301450
/organism="Clostridium perfringens str. 13"
/mol_type="Genomic DNA"
/strain="13"
/db_xref="taxon:195102"
/notes="anaerobic pathogen for gas gangrene"
64..342
/gene="CPE0728"
64..342
/gene="CPE0728"
/notes="92 aa, similar to pir:140869 hypothetical protein 4
nanH region from Clostridium perfringens (190 aa); 97.6%
identity in 84 aa overlap. Truncated by frameshift
mutation (confirmed by PCR-direct sequencing)."
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB80434.1"
gene
CDS
/db_xref="GI:18144388"
/translation="MNKKVILITGASSCIGYSTAKLLIEGHIYVCTARRVNNRELO
RLGKLIIFLDVKDERAVQKVNTTIRMESRIDVVVFANSWGIRNCKYR"
290..790
/gene="CPE0729"
290..790
/gene="CPE0729"
/notes="166 aa, similar to pir:140869 hypothetical protein
4 nanH region from Clostridium perfringens (190 aa); 97.1%
identity in 103 aa overlap. Truncated by frameshift
mutation (confirmed by PCR-direct sequencing)"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB80435.1"
/db_xref="GI:18144389"
/translation="MWFLTLTGGALETNIDDAKEEFDNLFVGVARIVKAVMTQMRKQ
KYGIIVTSSLAGIVSTPMSPFYPSKHALEGMIDGFRMENKESGIKIVKIQSPFINT
EPIKPFMNSLNKASKTKNAKAYKKEFEFFRKSSTKLILSSPPPEKVAKVSIIISKN
PKNSIL"
384..1448
/gene="CPE0730"
384..1448
/gene="CPE0730"
/notes="154 aa, similar to sp:YIAC.ECOLI HYPOTHETICAL 17.1
KDA PROTEIN IN TAG-BISC INTERGENIC REGION. from
Escherichia coli (146 aa); 36% identity in 139 aa overlap"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB80436.1"
/db_xref="GI:18144390"
/translation="MGERVLEIRBPKNCEIDDIMKWLSTVEAHYFIEEYWKKNYE
WRDIVIPAKTEFVCDGSKINGPTSIIDSNFIALFVHTKSOGSGIKSLLEYVKNK
YENIELAYKONKKXAVEYKXKHDFAKILKEQENEDSGHLEYLMSYKSSSENQ"
1804..1980
/gene="CPE0731"
1804..1980
/gene="CPE0731"
/notes="58 aa, no significant homology."
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB80437.1"
/db_xref="GI:18144391"
/translation="MKKCLYCKDLEKEPKESYIENKVGYFCSEDHFDKYILSTPEE
VIEVQNSFCVCSDD"
2333..2808
/gene="CPE0732"
2333..2808
/gene="CPE0732"
/notes="191 aa, similar to pir:H69339 conserved
hypothetical protein AF0720 from Archaeoglobus fulgidus
(214 aa); 38.5% identity in 213 aa overlap"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB80438.1"
/db_xref="GI:18144392"
/translation="MNEEDGKRSRAHSLFPSPVLAQAFAIGLPLFTAATNWDYKKNF
VKKLKEADLGAELVITGIDVPEEDCWHRVTSNIGLGLMPLMRKNHKEVVEEFIN
LGFVTKIVTNLKNMGKXEDLGRILITFDYIKELBERGIDPCGEAFETTTVIGGSLFK
KELRVNKGELISEGNMYLTLELDDFKLSDE"
complement(3023..3919)
/gene="CPE0733"
complement(3023..3919)
/gene="CPE0733"
/notes="298 aa, similar to sp:YCG30 ODOSI HYPOTHETICAL
TRANSCRIPTIONAL REGULATOR YCF30 from Odontella sinensis
chloroplast (309 aa); 29% identity in 262 aa overlap
LysR family"
/codon_start=1
/transl_table=11
```

```

/product="probable transcriptional regulator"
/protein_id="BAB80439.1"
/db_xref="GI:18144393"
/translation="MNLTLYLRSFYTTVKCNISISKAQQLHLTPQGVSMQIOKLENDIN
FKLNRSTNGVLSAGGIIIEFAEBSMISIDNLOKLDLKNISKLIIISCCSKSLGE
HVMPCSYTFEIKHYKTDISIDNIDNTLTKLTNHTNIGIIIOKPSDCNKLNLPL
MTDNLVVGNGRTSKSVNVELYKPLITREDGSANKLLVSSLLKKNKINIDNLNFV
LSLNSPQISKSVSFGQYSGLPALSLNHLRSSELKIDINDLTLPQYVIALRKNY
KLDYEQFVPLTSKKRCFCY"
4269..5969
/gene="naox"
/gene="naox"
/gene="naox"
/feature="566 aa, similar to gnu:AE004394.5 NADH oxidase,
putative from Vibrio cholerae (567 aa); 43% identity in
530 aa overlap. Putative N-terminal signal sequence was
found by PSORT
CPE0734"
/codon_start=1
/transl_table=11
/product="NADH oxidase"
/protein_id="BAB80440.1"
/db_xref="GI:18144394"
/translation="MSKIVIVGVGAGASTAARLRLENNOIMPERGPHVSFNC
CLPYLSKGVENHEDLVMTPEKFSQYRIDARVYTEVYSINRAKEVTVNVSQGE
YTESDKLVSPGAKAIVPPIKGIEDVNVFTVRNVDIADKLSFIKONNSKNI5VIGG
FGIYGEVAENHEAGYNTLTLEAMDOIIMKPDYDMVOVLHKEMVDKGNVLIVGDKVSS
PESGKVLSESKINDAVNAIGVAPETDLAREAGLEIGOTGAIKKNQYLNKDKLI
YAVGDVAVYNSLINSISRLPLAGPAQCAQARAVADHHRSPRSNSGVISSCVQIFHY
NGASTLGEQGLKATLNSINDVVRVIFDQKLVGMPGSEPLHFKLIIEVPFGRILGAG
AIGNADKRVDIITNLKMGTVEDDKLCEVAPPEFTAGDVNNAALDVASNLLEG
TFQYVHEVSDVESGACIIDVRESEYEAGHKGAKNPLSEIRDLRDIPTDRPV
"
YLCHRSQARSYNACLALQHLGFDNIYNSVGGFMGISFVEYVYNDKVQDREPITYNEN
"
6075..7361
/gene="CPE0735"
/feature="428 aa, partially similar to pir:D82968
hypothetical protein PA5430 from Pseudomonas aeruginosa
(strain PAO1) (404 aa); 25.5% identity in 314 aa overlap.
Putative N-terminal signal sequence and 7 putative
transmembrane regions were found by PSORT."
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB80441.1"
/db_xref="GI:18144395"
/translation="WKXTEYIIGFILLAAIIIGKINLGSSMLFRLLAGLGLGYAL
TRSFPGAGVNRSTKSTGKLMRLMILFVGTAVSVCFPIGQDPREYDLWNPINL
GLIIGLLFGFMFSSCCASGVLTDVVTGLPRAFITLIPFGMGVYLPVQNSASVY
KDTWISTGTFNGVFLPLDKWDLGELGAIIVITILACTIIVAKLYBEKRKANVT
YTGVDTEILOBKKEHEKEGKFLFSEATVEKLFVRPWTLATGALVIGGIFILLMGVTKA
GWGASTPYNFGALVAVGVSPEVSATFAGKPVEMPTAPFSGNMKNVQNGIILGL
VALLAGSKFTFKAGLQISEKALLIYVGGFANGFGTSLNGCNVCAITYPIANLSL
SGNPLFLVAVAGVIGNIYVYKMGKCKKEINNK"
7450..8781
/gene="CPE0736"
/feature="443 aa, similar to gnu:AP001510.169 BH0994 gene
product from Bacillus halodurans (395 aa); 24.3% identity
in 369 aa overlap. Putative N-terminal signal sequence and
11 putative transmembrane regions were found by PSORT."
/codon_start=1
/transl_table=11
/product="probable transporter"
/protein_id="BAB80442.1"
/db_xref="GI:18144396"
/translation="NKEIKLSRIDIFSIVLGAILGWGSMPLPGTKFLKESGVINTFUG
LFLGVCIIIEIKYLLINMQTHDEBGEFSFTYNNKLGKHGFIIVGFWLTLAYFTMPL
NATAPFLVIKFGGILFEGYVNAVAGNIYVGLIELTSSIIIIAFAALNGLKTKSK
VQNLIIFSLIGMTFIVLVGMIKGNRIDFINTYINKYSFDSLQIKKVPFATPPAFIFG

```

```

DAIPIQSKELNFSKKASRVAVISLVGALIVNLIITAIAYSPQASLSLEWAQSA
VLSILGRGAFFLLIIALTAAVWSGNGFMFICSKLGSIANYKMLPSRMKNKNGVF
SNAIITIVISLIAPFGRQAIIVIMSSLSGASVAYFYVSVILKESNTKDKILAG
IGVIGSIIFMLLILIPISPAALSKESLIALIWCITGFIAYYKIQDNIEA"
complement (8940..9602)
/gene="CPE0737"
complement (8940..9602)
/gene="CPE0737"
/feature="220 aa, similar to >gp:LM0132543.1
Query Match 18.4%; Score 40; DB 1; Length 301450;
Best Local Similarity 52.4%; Pred. No. 3;
Matches 88; Conservative 0; Mismatches 80; Indels 0; Gaps 0;
Qy 2 ATTCCGACGATGATGATACAGCGGATTAATACAGGTAAGTTCATTACACGGAAGAAC 61
Db 214171 ATTCAATATAATTTACTATAGATTATTAATAAAGGTGAATTTATGAAGATAGAAGC 214230
Qy 62 TTTAAGAATAATATGAAGGTATTACCTTGTGTTCACCGGAATGGTAAATATGCCCG 121
Db 214231 TTCAAGTACCATAGTAGAGTTCAAGAGTATTTCACCAAGTTCATTAATTCAGAGG 214290
Qy 122 AGTTTTCATGATACGGAATCCAGCAATTCCTTATTCATTAATTTGACT 169
Db 214291 AGTTATAGACTATAAACTTCCTTTATGTGATTTAATGAGTAGTATCT 214338
RESULT 14
AF412737/c AF412737 695 bp DNA linear INV 26-SEP-2001
LOCUS Inachis io NADH dehydrogenase 1 (ND1) gene, partial cds;
DEFINITION mitochondrial gene for mitochondrial product.
ACCESSION AF412737
VERSION AF412737.1 GI:15778377
KEYWORDS mitochondrion Inachis io
SOURCE Inachis io
ORGANISM Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
Papilionoidea; Nymphalidae; Nymphalinae; Inachis.
REFERENCE 1 (bases 1 to 695)
AUTHORS Nylin,S., Nyblom,K., Ronquist,P., Janz,N., Belicek,J. and
Kallersjo,M.
Phylogeny of Polytonia, Nymphalis and related butterflies
(Lepidoptera: Nymphalidae): a total-evidence analysis
JOURNAL Zool. J. Linn. Soc. 132, 441-468 (2001)
REFERENCE 2 (bases 1 to 695)
AUTHORS Nylin,S., Nyblom,K., Ronquist,P., Janz,N., Belicek,J. and
Kallersjo,M.
Direct Submission
TITLE Submitted (22-AUG-2001) Department of Zoology, Stockholm
JOURNAL University, Stockholm 106 91, Sweden
FEATURES Location/Qualifiers
source
1..695
/organism="Inachis io"
/organelle="mitochondrion"
/mol_type="genomic DNA"
/db_xref="taxon:171585"
<1..>695
/gene="ND1"
<1..>695
/gene="ND1"
/codon_start=1
/transl_table=5
/product="NADH dehydrogenase 1"
/protein_id="AA07408.1"
/db_xref="GI:15778378"
/translation="MIDMENIMIGLILIVGIVGAVTILERKVLGIQIRKGNK
MGVGLLQPCDAIKLFSKEQVLYNYSVLYFIPDINGFILSLXIMVYPMYFNMX
FNGLIFFXFFXCTQVLYTMAGSSNSNYSLGLGRAVQTIYSYEVMTLIMSSI
IIMDFNLKFNQYGLVWFIIMWMLPSFCFLSSLMAETNRTPYDFAEGESELVSGFN
IEYSSGCPALIFLA"
BASE COUNT 216 a 54 c 99 g 316 t 10 others

```





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 14, 2003, 23:55:19 ; Search time 140.052 Seconds  
(without alignments)  
4182.570 Million cell updates/sec

Title: US-09-928-457-99

Perfect score: 217

Sequence: 1 RAATCGGACATATCAATAC.....ATGACGATCAAGATTTAATT 217

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseq 19Jun03.\*  
1: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*  
2: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*  
3: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*  
4: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*  
5: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*  
6: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*  
7: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*  
8: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*  
9: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*  
10: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*  
11: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*  
12: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*  
13: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*  
14: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*  
15: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*  
16: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*  
17: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*  
18: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*  
19: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*  
20: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*  
21: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*  
22: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*  
23: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*  
24: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*  
25: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	217	100.0	217	AAV30599	Neisseria meningitidis
2	130.4	60.1	858	AAAI5296	DNA encoding a pol
3	130.4	60.1	858	AAAI5322	DNA encoding a pol
4	128.4	59.2	855	ABZ37755	N. gonorrhoeae nuc
5	128.4	59.2	891	ABZ41353	N. gonorrhoeae nuc
6	37.6	17.3	472	AEQ58421	Human colon cancer
7	37.6	17.3	602	ABK55260	Human colon cancer
8	37.6	17.3	2998	AAI18276	Lung cancer associ

C 9	37.6	17.3	3817	22	AAH02874	Human shear stress
C 10	37.6	17.3	3826	24	ABK94707	Human cDNA differe
C 11	36	16.6	2000	24	ABZ15907	Arabidopsis thalia
C 12	36	16.6	640681	24	ABK92787	Buchnera sp. genom
C 13	35.4	16.3	163350	24	AAK46127	Human tumour suppr
C 14	34.8	16.0	17421	22	AAK45349	Chemically pretrea
C 15	34.8	16.0	17421	24	ABK28182	DNA transcription
C 16	34.4	15.9	339	22	ABAI2692	Human nervous syst
C 17	34.4	15.9	393	22	ABAI1810	Human nervous syst
C 18	34.4	15.9	393	22	ABAI1811	Human nervous syst
C 19	34.4	15.9	393	22	ABAI1812	Human nervous syst
C 20	34.2	15.8	2047	17	AAI14767	Campylobacter jeju
C 21	33.8	15.6	584	23	ABK52653	Human prostate exp
C 22	33.6	15.5	540	24	ABK60957	Human cancer relat
C 23	33.6	15.5	4873	22	AAK81633	Human immune/haema
C 24	33.2	15.3	1498	23	ABV20826	Human prostate exp
C 25	33.2	15.3	1498	23	ABV21352	Human prostate exp
C 26	33.2	15.3	1498	23	ABV21414	Human prostate exp
C 27	33.2	15.3	1498	23	ABV21491	Human prostate exp
C 28	33.2	15.3	1498	23	ABV21495	Human prostate exp
C 29	33.2	15.3	1498	23	ABV21507	Human prostate exp
C 30	33.2	15.3	1498	23	ABV21565	Human prostate exp
C 31	33.2	15.3	1498	23	ABV21675	Human prostate exp
C 32	33.2	15.3	1498	23	ABV21700	Human prostate exp
C 33	33.2	15.3	1498	23	ABV26673	Human prostate exp
C 34	33.2	15.3	1498	23	ABV27171	Human prostate exp
C 35	33.2	15.3	1498	23	ABV27232	Human prostate exp
C 36	33.2	15.3	1498	23	ABV27309	Human prostate exp
C 37	33.2	15.3	1498	23	ABV27313	Human prostate exp
C 38	33.2	15.3	1498	23	ABV27325	Human prostate exp
C 39	33.2	15.3	1498	23	ABV27384	Human prostate exp
C 40	33.2	15.3	1498	23	ABV27496	Human prostate exp
C 41	33.2	15.3	1498	23	ABV27522	Human prostate exp
C 42	33	15.2	334	19	AAV03336	DNA sequence that
C 43	32.8	15.1	18434	24	ABK34006	Human immune syste
C 44	32.6	15.0	301	25	ABK33229	Human chromosome 3
C 45	32.6	15.0	461	22	AAH13064	Human cDNA clone (

## ALIGNMENTS

RESULT 1  
AAV30599  
ID AAV30599 standard; DNA; 217 BP.  
XX  
AC AAV30599;  
XX  
DT 22-OCT-1998 (first entry)  
XX  
DE Neisseria meningitidis DNA sequence E110.  
XX  
KW N. Gonorrhoeae; N. lactamica; chromosome 22491; region 1; region 2;  
KW region 3; pathogenicity; blood-brain barrier; diagnosis; infection;  
KW meningitis; ss.  
XX  
OS Neisseria meningitidis.  
XX  
PN WO9802547-A2.  
XX  
PD 22-JAN-1998.  
DF 11-JUL-1997; 97WO-FR01295.  
XX  
PR 12-JUL-1996; 96FR-0008768.  
XX  
PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.  
PA (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.  
XX (SMIX ) SMITHKLINE BEECHAM.  
XX  
PI Nassif X, Tinsley C, Achtman M, Merker P, Ruelle J;  
PI Vinals C;  
XX

DR WPI; 1998-110594/10.

XX Genes present in *Neisseria meningitidis* but not other *Neisseria*

PT species - and related host cells, RNA, anti-sense sequences,

PT polypeptide(s) and antibodies, useful for diagnosing *Neisseria*

PT meningitidis infection and in protective vaccines

XX

PS Example 4; Page 131; 150pp; French.

XX This sequence represents a sequence present in *Neisseria meningitidis*

CC and *N. gonorrhoeae* but not in *N. lactamica*, except for the genes involved

CC in biosynthesis of the capsule polysaccharide, frpA or C, opc, porA,

CC rotamase, sequence IC1106, IGA protease, pillin, pilC, proteins which

CC bind transferrin and opacity proteins. The DNA sequences are responsible

CC for the differences in pathogenicity between *N. meningitidis* and

CC *N. gonorrhoeae*, specifically they include the genes that allow

CC *N. meningitidis* to cross the blood-brain barrier. DNA sequences common

CC to *N. meningitidis* and *N. gonorrhoeae*, but absent from *N. lactamica*,

CC are responsible for colonisation and penetration of the mucosa. The

CC DNA sequences can be used to produce probes and primers, and

CC antibodies produced against the encoded proteins are used in standard

CC hybridisation/immunoassay processes for diagnosis of *N. meningitidis*

CC infection, particularly meningitis.

XX

SQ Sequence 217 BP; 76 A; 31 C; 44 G; 66 T; 0 other;

Query Match 100.0%; Score 217; DB 19; Length 217;

Best Local Similarity 100.0%; Pred. No. 6.2e-53;

Matches 217; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AATCGGCAGTATGATACAGCGGATTAATACAGGTAGTTTCAATACACGGAAAAAC 60

DB 1 AATCGGCAGTATGATACAGCGGATTAATACAGGTAGTTTCAATACACGGAAAAAC 60

QY 61 CTTTAAAGAAATATGAAAGGTATTACCTTGTGGCAACGGGAATGGTAATATGCC 120

DB 61 CTTTAAAGAAATATGAAAGGTATTACCTTGTGGCAACGGGAATGGTAATATGCC 120

QY 121 GAGTTTTTCATGAAATACCGGAATCCAGCAATTTCTATTCATATTGACTGGTGA 180

DB 121 GAGTTTTTCATGAAATACCGGAATCCAGCAATTTCTATTCATATTGACTGGTGA 180

QY 181 TGTGCACATTTAGATATGATGAGATGAGATTAATT 217

DB 181 TGTGCACATTTAGATATGATGAGATGAGATTAATT 217

RESULT 2

AAAL5296

ID AAA15296 standard; DNA; 858 BP.

XX

AC AAA15296;

XX

DT 04-SEP-2000 (first entry)

XX

DE DNA encoding a polypeptide of a *Neisseria* pathogenic strain.

XX

KW Pathogenic strain; *Neisseria*; vaccine; *Neisseria* infection; ss.

XX

OS *Neisseria meningitidis*.

XX

PH Key Location/Qualifiers

FT CDS 1..858

FT /\*tag= a

XX

PN WO200026375-A2.

XX

XX 11-MAY-2000.

XX

PD 28-OCT-1999; 99WO-FR02643.

XX

PF 30-OCT-1998; 98FR-0013693.

XX

PA (INMR) PASTEUR MERIEUX SERUMS & VACCINS SA.

XX (INRM) INSERM INST NAT SANTE & RECH MEDICALE.

XX Aujame L, Bouchardon A, Renauld-Mongenie G, Rokbi B, Nassif X;

XX Tinsley C, Perrin A;

PA (INMR) PASTEUR MERIEUX SERUMS & VACCINS SA.

PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.

XX Aujame L, Bouchardon A, Renauld-Mongenie G, Rokbi B, Nassif X;

PI Tinsley C, Perrin A;

XX WPI; 2000-365622/31.

DR P-PSDB; AAY93265.

XX

PT New polypeptide specific for pathogenic *Neisseria* useful in therapeutic

PT or preventative vaccines and for diagnosis

XX

PS Claim 2; Page 51-52; 187pp; French.

XX The present sequence encodes a protein that is specific for pathogenic

CC strains of *Neisseria*. The polynucleotides, polypeptides, or their

CC antigenic fragments, are used in vaccines to treat or protect against

CC *Neisseria* infections, particularly by *N. meningitidis*. The

CC polynucleotide sequence is also used for recombinant production of

CC the polypeptide and to produce attenuated *Neisseria* strains that

CC overexpress it, or express it in a non-toxic mutant form.

XX

SQ Sequence 858 BP; 297 A; 145 C; 182 G; 234 T; 0 other;

Query Match 60.1%; Score 130.4; DB 21; Length 858;

Best Local Similarity 97.5%; Pred. No. 6.7e-28;

Matches 154; Conservative 0; Mismatches 1; Indels 3; Gaps 2;

QY 1 AATCGGCAGTATGATACAGCGGATTAATACAGGTAGTTTCAATACACGGAAAAAC 60

DB 701 AATCGGCAGTATGATACAGCGGATTAATACAGGTAGTTTCAATACACGGAAAAAC 760

QY 61 CTTTAAAGAAATATGAAAGGTATTACCTTGTGGCAACGG--GAATGGTAATATGC 118

DB 761 CTTTAAAGAAATATGAAAGGTATTACCTTGTGGCAACGGAGATGGTAATATGC 820

QY 119 CCGAGTTTTTCACTGAAATAGCGAATCCAGCCATTCTA 156

DB 821 CCGAGTTTTTCACTG-ATAGCGAATCCAGCCATTCTA 857

RESULT 3

AAAL5322

ID AAA15322 standard; DNA; 858 BP.

XX

AC AAA15322;

XX

DT 04-SEP-2000 (first entry)

XX

DE DNA encoding a polypeptide of a *Neisseria* pathogenic strain.

XX

KW Pathogenic strain; *Neisseria*; vaccine; *Neisseria* infection; ss.

XX

OS *Neisseria gonorrhoeae*.

XX

PH Key Location/Qualifiers

FT CDS 1..858

FT /\*tag= a

XX

PN WO200026375-A2.

XX

XX 11-MAY-2000.

XX

PD 28-OCT-1999; 99WO-FR02643.

XX

PF 30-OCT-1998; 98FR-0013693.

XX

PA (INMR) PASTEUR MERIEUX SERUMS & VACCINS SA.

XX (INRM) INSERM INST NAT SANTE & RECH MEDICALE.

XX Aujame L, Bouchardon A, Renauld-Mongenie G, Rokbi B, Nassif X;

XX Tinsley C, Perrin A;

DR WPI; 2000-365622/31.  
 DR P-PSDB; AAY93292.  
 XX  
 PT New polypeptide specific for pathogenic Neisseria useful in therapeutic  
 PT or preventative vaccines and for diagnosis -  
 XX  
 PS Claim 4; Page 134-135; 187pp; French.  
 XX  
 CC The present sequence encodes a protein that is specific for pathogenic  
 CC strains of Neisseria. The polynucleotides, polypeptides, or their  
 CC antigenic fragments, are used in vaccines to treat or protect against  
 CC Neisseria infections, particularly by N. meningitidis. The  
 CC polynucleotide sequence is also used for recombinant production of  
 CC the polypeptide and to produce attenuated Neisseria strains that  
 CC overexpress it, or express it in a non-toxic mutant form.  
 XX  
 SQ Sequence 858 BP; 302 A; 151 C; 178 G; 227 T; 0 other;

Query Match 60.1%; Score 130.4; DB 21; Length 858;  
 Best Local Similarity 97.5%; Pred. No. 6.7e-28;  
 Matches 154; Conservative 0; Mismatches 1; Indels 3; Gaps 2;

QY 1 AATTGGACAGTATGAATACAGCGGATTAAATACAGGTAAGTTTCATTACACGGGAARAC 60  
 Db 701 AATTGCAACAGTATGAATACAGCGGATTAAATACAGGTAAGTTTCATTACACGGGAARAC 760

QY 61 CTTTAAAGAATAATGAAGAAGTATTACCTTTTGGCCAAACGG--GAATGTTAAATATGC 118  
 Db 761 CTTTAAAGAATAATGAAGAAGTATTACCTTTTGGCCAAACGGGAATGTTAAATATGC 820

QY 119 CCGAGTTTTTCACTGAATAGCGAATCCAGCCATTCTA 156  
 Db 821 CCGAGTTTTTCACTG-ATAGCGAATCCAGCCATTCTA 857

RESULT 4  
 ABZ37755  
 ID ABZ37755 standard; DNA; 855 BP.  
 XX  
 AC ABZ37755;  
 XX  
 DT 07-MAR-2003 (first entry)  
 XX  
 DE N. gonorrhoeae nucleotide sequence SEQ ID 99.  
 XX  
 KW Antibacterial; infection; vaccine; gene therapy; gene; ds.  
 XX  
 OS Neisseria gonorrhoeae.  
 XX  
 PN WO200279243-A2.  
 XX  
 PD 10-OCT-2002.  
 XX  
 PF 12-FEB-2002; 2002WO-IB02069.  
 XX  
 PR 12-FEB-2001; 2001GB-0003424.  
 XX  
 PS (CHIR-) CHIRON SPA.  
 XX  
 PI Fontana MR, Pizza M, Masignani V, Monaci E;  
 XX  
 PN WPI; 2003-058415/05.  
 XX  
 PD P-PSDB; ABP76785.  
 XX  
 PF New protein from Neisseria gonorrhoeae, useful for the manufacture of a  
 PT medicament for treating or preventing N. gonorrhoeae infection -  
 XX  
 PS Claim 6; Page 188; 815pp; English.  
 XX  
 CC The present invention relates to proteins from Neisseria gonorrhoeae.  
 CC Also disclosed are the nucleic acid molecules encoding the proteins and  
 CC antibodies that specifically bind to the proteins. The composition  
 CC comprising the protein, nucleic acid or antibody is useful for the  
 CC manufacture of a medicament for treating or preventing N. gonorrhoeae  
 CC infection, this may be in the form of a vaccine or gene therapy.  
 CC Sequences given in records ABZ37706-ABZ42016 represent nucleic acid  
 CC molecules of the invention.  
 XX  
 SQ Sequence 891 BP; 316 A; 153 C; 186 G; 236 T; 0 other;

Query Match 59.2%; Score 128.4; DB 25; Length 891;  
 Best Local Similarity 97.4%; Pred. No. 2.6e-27;  
 Matches 152; Conservative 0; Mismatches 1; Indels 3; Gaps 2;

QY 1 AATTGGACAGTATGAATACAGCGGATTAAATACAGGTAAGTTTCATTACACGGGAARAC 60  
 Db 701 AATTGCAACAGTATGAATACAGCGGATTAAATACAGGTAAGTTTCATTACACGGGAARAC 760

QY 61 CTTTAAAGAATAATGAAGAAGTATTACCTTTTGGCCAAACGG--GAATGTTAAATATGC 118  
 Db 761 CTTTAAAGAATAATGAAGAAGTATTACCTTTTGGCCAAACGGGAATGTTAAATATGC 820

QY 119 CCGAGTTTTTCACTGAATAGCGAATCCAGCCATTTC 154  
 Db 821 CCGAGTTTTTCACTG-ATAGCGAATCCAGCCATTTC 855

RESULT 5  
 ABZ41353  
 ID ABZ41353 standard; DNA; 891 BP.  
 XX  
 AC ABZ41353;  
 XX  
 DT 07-MAR-2003 (first entry)  
 XX  
 DE N. gonorrhoeae nucleotide sequence SEQ ID 7295.  
 XX  
 KW Antibacterial; infection; vaccine; gene therapy; gene; ds.  
 XX  
 OS Neisseria gonorrhoeae.  
 XX  
 PN WO200279243-A2.  
 XX  
 PD 10-OCT-2002.  
 XX  
 PF 12-FEB-2002; 2002WO-IB02069.  
 XX  
 PR 12-FEB-2001; 2001GB-0003424.  
 XX  
 PS (CHIR-) CHIRON SPA.  
 XX  
 PI Fontana MR, Pizza M, Masignani V, Monaci E;  
 XX  
 PN WPI; 2003-058415/05.  
 XX  
 PD P-PSDB; ABP80383.  
 XX  
 PF New protein from Neisseria gonorrhoeae, useful for the manufacture of a  
 PT medicament for treating or preventing N. gonorrhoeae infection -  
 XX  
 PS Disclosure; Page 718; 815pp; English.  
 XX  
 CC The present invention relates to proteins from Neisseria gonorrhoeae.  
 CC Also disclosed are the nucleic acid molecules encoding the proteins and  
 CC antibodies that specifically bind to the proteins. The composition  
 CC comprising the protein, nucleic acid or antibody is useful for the  
 CC manufacture of a medicament for treating or preventing N. gonorrhoeae  
 CC infection, this may be in the form of a vaccine or gene therapy.  
 CC Sequences given in records ABZ37706-ABZ42016 represent nucleic acid  
 CC molecules of the invention.  
 XX  
 SQ Sequence 891 BP; 316 A; 153 C; 186 G; 236 T; 0 other;

Query Match 59.2%; Score 128.4; DB 25; Length 891;  
 Best Local Similarity 97.4%; Pred. No. 2.6e-27;  
 Matches 152; Conservative 0; Mismatches 1; Indels 3; Gaps 2;

QY 1 AATTCGGACATGATACACAGCGGATTATACAGGTAAGTTTCATTACACGGAATAAC 60  
 DB 737 AATTCGACAGATGATACACAGCGGATTATACAGGTAAGTTTCATTACACGGAATAAC 796  
 QY 61 CTTTAAAGATAAATGATGAAGGTTATACCTTGTGTCACAGG--GAATGTTAAATATGC 118  
 DB 797 CTTTAAAGATAAATGATGAAGGTTATACCTTGTGTCACAGGCGAATGTTAAATATGC 856  
 QY 119 CCGAGTTTTCCTGTAAGTACGGAATCCAGCCATTTC 154  
 DB 857 CCGAGTTTTCCTG-ATAGCGAATCCAGCCATTTC 891

## RESULT 6

ABQ58421  
 ID ABQ58421 standard; cDNA; 472 BP.

XX AC ABQ58421;

XX DT 02-AUG-2002 (first entry)

XX DE Human colon cancer related nucleotide sequence SEQ ID NO:2116.

XX DE Human; colon cancer; cancer; tissue profiling; forensic; mapping;  
 KW genetic analysis; diagnostic; antisense therapy; gene; ss.

XX OS Homo sapiens.

XX PN WO200229086-A2.

XX PD 11-APR-2002.

XX PF 02-OCT-2001; 2001WO-US30732.

XX PR 02-OCT-2000; 2000US-237271P.

XX PA (FARB ) BAYER CORP.

XX PI Burgess C, Astle JH, Carroll E, Catino TJ, Dwivedi P, Molino GA;  
 PI Thiagalingam A, Lewis ME;

XX DR WPI; 2002-426115/45.

XX PT New isolated nucleic acid that is differentially expressed in cancer  
 PT tissues useful for determining the presence of colon cancer in a cell  
 PT or tissue type, and in antisense therapy

XX PS Claim 1; Fig 1; 796pp; English.

XX CC ABQ56306 to ABQ60787 represent isolated nucleic acids (I) differentially  
 CC expressed in cancer tissues. ABQ78993 to ABQ79004 represent proteins  
 CC encoded by the ABQ60776 to ABQ60787 nucleic acid sequences. (I) can be  
 CC used in antisense therapy. An antibody immunoreactive with a polypeptide  
 CC encoded by (I) is useful for detecting cancer in a patient sample, and  
 CC for detecting the presence or absence of a polynucleotide encoded by a  
 CC nucleic acid which hybridizes to (I) in a cell. A probe/primer derived  
 CC from (I) can be used for determining the presence of a nucleic acid which  
 CC hybridizes to (I), and for determining the phenotype of cells in a sample  
 CC of cells from a patient. (I) is useful for determining the presence of  
 CC colon cancer in a cell or tissue type, for determining the presence or  
 CC state of other type of cancer, in antisense therapy, to generate  
 CC macroarrays on a solid surface, to identify a chromosome on which the  
 CC corresponding gene resides, and in tissue profiling, forensics, genetic  
 CC analysis, mapping and diagnostic applications. (I) can be used to raise  
 CC antibodies, and to screen for peptide analogues and antagonists.

XX SQ Sequence 472 BP; 172 A; 100 C; 60 G; 132 T; 8 other;

Query Match 17.3%; Score 37.6; DB 24; Length 472;  
 Best Local Similarity 51.8%; Pred. No. 0.29;  
 Matches 85; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 1 AATTCGGACATGATGAATACAGCGGATTATACAGGTAAGTTTCATTACACGGAATAAC 60

DB 17 AACTGAGGCATTAAATACCTCTTTATAATAATGCAAGTTGAAATGCTAACAAAGCATAAAC 76  
 QY 61 CTTTAAAGATAAATGATGAAGGTTATACCTTGTGTCACAGGGAATGTTAAATATGCC 120  
 DB 77 ACTTCTGCAAAAATTCACAGGACAGTTGTTTCAATCAAGAGAAAAGTCATAACCACT 136  
 QY 121 GAGTTTTTCACTGAATAGCAATCCAGCCATTTCTTATTCATATT 164  
 DB 137 TGGTTTTTAAATGAAAATCCCTTCACATCCACCTGTGTTCAAAAT 180

## RESULT 7

ABK55260/c

ID ABK55260 standard; cDNA; 602 BP.

XX AC ABK55260;

XX DT 18-JUN-2002 (first entry)

XX DE Human colon cancer-associated cDNA, SEQ ID NO 730.

XX DE Human; colon cancer; immunogenic; vaccine; tumour; gene; ss.

XX OS Homo sapiens.

XX PN WO200212280-A2.

XX PD 14-FEB-2002.

XX PF 30-JUL-2001; 2001WO-US23826.

XX PR 03-AUG-2000; 2000US-223265P.

XX PR 02-OCT-2000; 2000US-237406P.

XX PR 20-MAR-2001; 2001US-277495P.

XX PR 03-JUL-2001; 2001US-302702P.

XX PA (CORI-) CORIXA CORP.

XX PI Pyle RA, Xu J, Secrist H;

XX DR WPI; 2002-257462/30.

XX PT Novel polynucleotide encoding colon tumour polypeptides, useful as  
 PT vaccines for treating colon cancers

XX PS Claim 1; Page 356; 425pp; English.

XX CC The invention relates to isolated polynucleotides (I) encoding colon  
 CC tumour polypeptides (II). (I) is useful for stimulating an immune  
 CC response in a patient and treating colon cancer in a patient.  
 CC Oligonucleotides derived from (I) are useful for determining the presence  
 CC of cancer in a patient. (I) and (II) are useful in pharmaceutical  
 CC compositions, e.g. vaccines, and other compositions for the diagnosis  
 CC and treatment of colon cancer. A composition comprising a first component  
 CC selected from physiologically acceptable carriers and immunostimulants,  
 CC and an antigen-presenting cell expressing (II) is useful for inhibiting  
 CC development of cancer in a patient. (I) is useful in the design and  
 CC preparation of ribozyme molecules for inhibiting expression of tumour  
 CC polypeptides and (I). ABK54531-ABK55464 represent human colon cancer cDNA  
 CC sequences of the invention.

XX SQ Sequence 602 BP; 155 A; 92 C; 120 G; 235 T; 0 other;

Query Match 17.3%; Score 37.6; DB 24; Length 602;  
 Best Local Similarity 51.8%; Pred. No. 0.3;  
 Matches 85; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 1 AATTCGACAGTATGAATACAGCGGATTATACAGGTAAGTTTCATTACACGGAATAAC 60

DB 417 AACTGAGGCATTAAATACCTCTTTATAATAATGCAAGTTGAAATGCTAACAAAGCATAAAC 358

QY 61 CTTTAAAGATAAATGATGAAGGTTATACCTTGTGTCACAGGGAATGTTAAATATGCC 120

Db 357 ACTTCTGCAAAATTCACAGGCACAGTGTTCATTCAACAGAAAAAGTCAAAACCACT 298  
 QY 121 GAGTTTTCACGAATAGGAATCCAGGCATTTCATTATCATATT 164  
 Db 297 TGGTTTTTAAATGAAATCCTTCACATCCACCTGTGTTCAAAAT 254

RESULT 8  
 AAF18276/c  
 ID AAF18276 standard; DNA; 2998 BP.  
 XX  
 AC AAF18276;  
 XX  
 DT 14-MAR-2001 (first entry)  
 XX  
 DE Lung cancer associated polynucleotide sequence SEQ ID 295.  
 XX  
 DE Human; lung cancer associated protein; neuroprotective; cytostatic;  
 KW cardioactive; immunomodulatory; muscular active; vulnerary;  
 KW gastrointestinal; nephrotropic; anti-infective; gynecological;  
 KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;  
 KW proliferative disorder; wound healing; infectious disease; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200055180-A2.  
 XX  
 PD 21-SEP-2000.  
 XX  
 XX 08-MAR-2000; 2000WO-US05918.  
 XX  
 PR 12-MAR-1999; 99US-0124270.  
 XX  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA (ROSE/) ROSEN C A.  
 XX  
 PI Ruben SM;  
 XX  
 DR WPI; 2000-587514/55.  
 DR P-PSDB; AAB58400.  
 XX

Lung cancer associated gene sequences, referred to as lung cancer  
 PT antigens, useful for treatment, prevention, and diagnosis of disorders  
 PT such as lung cancer -  
 XX  
 PS Claim 1; Page 754-755; 1425pp; English.

Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer  
 CC associated proteins represented in AAB58106 - AAB58548. Lung cancer  
 CC associated proteins and polynucleotide sequences, their agonists, and  
 CC antagonists may have neuroprotective; cytostatic; cardioactive; and  
 CC immunomodulatory; muscular active general; vulnerary; gastrointestinal  
 CC general; nephrotropic; anti-infective; gynecological; or antibacterial  
 CC activity. The invention also includes antibodies specific for the  
 CC protein or polynucleotide sequences. The lung cancer associated  
 CC polynucleotide sequences may be used for detection of lung cancer,  
 CC chromosome identification, as chromosome markers, and for numerous other  
 CC diagnostic or research purposes. The proteins may be used to treat  
 CC disorders such as neural, immune, muscular, reproductive,  
 CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative  
 CC disorders. The proteins may also be used in the treatment of wounds and  
 CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and  
 CC peptide AAB58549 are used in the course of the invention for the  
 CC identification and characterisation of the polynucleotide and protein  
 CC sequences.

Sequence 2998 BP; 826 A; 568 C; 699 G; 906 T; 9 other;  
 SQ

Query Match 17.3%; Score 37.6; DB 21; Length 2998;  
 Best Local Similarity 51.8%; Pred. No. 0.43; Indels 0; Gaps 0;  
 Matches 85; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 1 AATTCGGACAGTATGAATACAGCGGATTAATACAAGGTAAAGTTCAATACACGGAATAAC 60  
 Db 2282 AACTGAGGCATTAAATACCTCTTTTATAATATGCAAGTTGAAATGCTAACAAAGCATAAAC 2223  
 QY 61 CTTTAAAGATAATATGAAGGTATTACCTTTGTTTGCACAGCGGAATGTAATATGCC 120  
 Db 2222 ACTTCTGCAAAATTCACAGGCACAGTGTTCATTCAACAGAAAAAGTCAAAACCACT 2163  
 QY 121 GAGTTTTCACCTGAATAGCGAATCCAGGCATTTCATTATCATATT 164  
 Db 2162 TGGTTTTTAAATGAAATCCTTCACATCCACCTGTGTTCAAAAT 2119

RESULT 9  
 AAH02874/c  
 ID AAH02874 standard; DNA; 3817 BP.  
 XX  
 AC AAH02874;  
 XX  
 DT 15-JUN-2001 (first entry)  
 XX  
 DE Human shear stress-response coding sequence SEQ ID NO: 1.  
 XX Human shear stress-response protein; vascular disease;  
 KW Human; shear stress-response protein; vascular disease;  
 KW arteriosclerosis; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200125427-A1.  
 XX  
 PD 12-APR-2001.  
 XX  
 XX 02-OCT-2000; 2000WO-JP06840.  
 PF  
 XX 01-OCT-1999; 99JP-0280976.  
 PR  
 XX (KYOW) KYOWA HAKKO KOGYO KK.  
 PA (NOJI) NOJIMA H.  
 XX  
 PI Nojima H, Yoshisue H, Obayashi M, Ota T, Kawabata A, Sakurada K;  
 FI Kuga T, Sekine S, Nakamura Y, Sugano S;  
 XX  
 DR WPI; 2001-266308/27.  
 DR P-PSDB; AAB90751.  
 XX

DNA sequences, proteins encoded by them and antibodies against them  
 PT useful in diagnosis and treatment of vascular disease caused by  
 PT arteriosclerosis -  
 XX  
 PS Claim 20; Page 144-149; 678pp; Japanese.

The present invention provides the protein and coding sequences of a  
 CC number of human shear stress response proteins. These are useful in the  
 CC diagnosis, treatment and screening of vascular diseases caused by  
 CC arteriosclerosis, including heart failure, post-PTCA restenosis and  
 CC hypertension.

Sequence 3817 BP; 1051 A; 736 C; 936 G; 1094 T; 0 other;  
 SQ

Query Match 17.3%; Score 37.6; DB 22; Length 3817;  
 Best Local Similarity 51.8%; Pred. No. 0.46; Indels 0; Gaps 0;  
 Matches 85; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 1 AATTCGGACAGTATGAATACAGCGGATTAATACAAGGTAAAGTTCAATACACGGAATAAC 60  
 Db 3162 AACTGAGGCATTAAATACCTCTTTTATAATGCAAGTTGAAATGCTAACAAAGCATAAAC 3103  
 QY 61 CTTTAAAGATAATATGAAGGTATTACCTTTGTTTGCACAGCGGAATGTAATATGCC 120  
 Db 3102 ACTTCTGCAAAATTCACAGGCACAGTGTTCATTCAACAGAAAAAGTCAAAACCACT 3043  
 QY 121 GAGTTTTCACCTGAATAGCGAATCCAGGCATTTCATTATCATATT 164

DB 3042 TGGTTTAAATGAATCCTTCACATCCACCTGTTCAAAAT 2999

RESULT 10  
ID ABK84707/c  
XX ABK84707 standard; cDNA; 3826 BP.  
AC ABK84707;  
XX  
XX 14-AUG-2002 (first entry)  
XX Human cDNA differentially expressed in granulocytic cells #1279.  
XX  
XX Human; ss: granulocytic cell; DNA chip; bacterial infection;  
KW viral infection; parasitic infection; protozoal infection;  
KW fungal infection; sterile inflammation; disease; psoriasis;  
KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;  
KW cardiac reperfusion injury; renal reperfusion injury; ARDS;  
KW adult respiratory distress syndrome; inflammatory bowel disease;  
KW Crohn's disease; ulcerative colitis; periodontal disease;  
KW granulocyte activation; chronic inflammation; allergy.  
XX  
XX Homo sapiens.  
XX WO200228999-A2.  
XX  
XX 11-APR-2002.  
XX  
XX 03-OCT-2001; 2001WO-US30821.  
XX  
XX 03-OCT-2000; 2000US-237189P.  
XX (GENE-) GENE LOGIC INC.  
XX  
XX Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;  
XX WPI; 2002-435328/46.  
XX  
XX Detecting granulocyte activation by detecting differential expression  
PT of genes associated with granulocyte activation, which serves as  
PT diagnostic markers that is useful for monitoring disease states and  
PT drug toxicity -  
XX  
XX Claim 1; SEQ ID No 1278; 114pp; English.  
XX  
XX The invention relates to detecting (M1) granulocyte (GC) activation  
CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by  
CC DNA chip analysis as given in the specification, and comparing  
CC the expression level to an expression level in an unactivated  
CC GC, where differential expression of Gs is indicative of GCA.  
CC Also included are modulating (M2) GA by contacting GC with an agent  
CC that alters the expression of at least one gene in Gs; (2) screening (M3)  
CC for an agent capable of modulating GCA or an inflammation (especially  
CC chronic) in a tissue, an allergic response in a subject, exposure of a  
CC subject to a pathogen or sterile inflammatory disease using the  
CC gene expression profile; (3) detecting (M4) an inflammation (especially  
CC chronic) in a tissue, an allergic response in a subject, exposure of a  
CC subject to a pathogen or sterile inflammatory disease, by detecting the  
CC level of expression in a sample of the tissue of gene(s) from Gs, where  
CC the level of expression of the gene is indicative of inflammation;  
CC (4) treating (M5) an inflammation (especially chronic) or in a tissue,  
CC an allergic response in a subject, exposure of a subject to a pathogen  
CC or sterile inflammatory disease, by contacting a tissue having  
CC inflammation with an agent that modulates the expression of gene(s)  
CC from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for  
CC modulating GCA; M3 is useful for screening an agent capable of modulating  
CC GCA preferably in an inflammation in a tissue; M4 is useful for  
CC detecting an inflammation (especially chronic) in a tissue, an allergic  
CC response in a subject, exposure of a subject to a pathogen or sterile  
CC inflammatory disease (e.g. psoriasis, rheumatoid arthritis,  
CC glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal  
CC reperfusion injury, ARDS, adult respiratory distress syndrome,  
CC inflammatory bowel disease, Crohn's disease, ulcerative colitis,

CC periodontal disease; also bacterial infection, viral infection,  
CC parasitic infection, protozoal infection, fungal infection and M5 is  
CC useful for treating one of the above conditions. The present  
CC sequence represents a gene differentially expressed in granulocytes.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 3826 BP; 1053 A; 740 C; 937 G; 1096 T; 0 other;  
XX  
XX Query Match 17.3%; Score 37.6; DB 24; Length 3826;  
XX Best Local Similarity 51.8%; Pred. NO. 0.46;  
XX Matches 85; Conservative 0; Mismatches 79; Indels 0; Gaps 0;  
XX  
QY 1 AATTCGGACAGTATGATATACAGCGGATTAAATACAGGTAAGTTCATTACACGGAAC 60  
DB 3162 AACTGAGGCAATTATACCTCTTTATTAATGCAAGTTGAATGCTACAAAGCATAAAC 3103  
QY 61 CTTTAAAGAATAATATGAAGGTATACCTTGTTCGCAACGGGAATGTTAAATATGCC 120  
DB 3102 ACTTCTGCAAAAATTCACAGGCACAGTTGTTTCATTCACAGAAAAAGTCAAAACCACT 3043  
QY 121 GAGTTTTCACGTAGTACGAATCCAGCCATTTCTATTCAATATT 164  
DB 3042 TGGTTTAAATGAATCCTTCACATCCACCTGTTCAAAAT 2999  
XX  
RESULT 11  
ABZ15907  
ID ABZ15907 standard; DNA; 2000 BP.  
XX AC ABZ15907;  
XX  
XX 21-JAN-2003 (first entry)  
XX  
XX Arabidopsis thaliana stress regulated gene SEQ ID NO 3712.  
XX  
XX Arabidopsis thaliana; plant; gene; stress; transgenic; ds.  
XX  
XX Arabidopsis thaliana.  
XX OS  
XX Arabidopsis thaliana.  
XX PN WO200216655-A2.  
XX  
XX 28-FEB-2002.  
XX  
XX 24-AUG-2001; 2001WO-US26685.  
XX  
XX 24-AUG-2000; 2000US-227866P.  
XX 26-JAN-2001; 2001US-264647P.  
XX 22-JUN-2001; 2001US-300111P.  
XX  
XX (SCRI ) SCRIPPS RES INST.  
XX (SYGN ) SYNGENTA PARTICIPATIONS AG.  
XX  
XX Harper JF, Kreps J, Wang X, Zhu T;  
XX WPI; 2002-304127/34.  
XX  
XX Identifying a stress condition to which a plant cell has been exposed  
PT and producing plants with increased tolerance to these abiotic stresses  
PT -  
XX  
XX Claim 144; SEQ ID NO 3712; 577pp + Sequence Listing; English.  
XX  
XX The invention relates to identifying a stress condition to which a plant  
CC cell has been exposed, comprising:  
CC (a) contacting nucleic acid representative of expressed polynucleotides  
CC in the plant cell with an array or probes representative of the plant  
CC cell genome; and  
CC (b) detecting a profile of expressed polynucleotides in the plant cell  
CC characteristic of a stress response. The method is useful in the  
CC production of transgenic plants, cells and seeds and in producing plants



CC preventing or treating inflammation, or disorders associated with cell  
CC proliferation and apoptosis e.g. bone cancer, brain cancer, cervix  
CC cancer, adenocarcinoma, leukaemia, lymphoma or melanoma. They are  
CC particularly useful as models for developing human therapeutic targets,  
CC identifying therapeutic proteins, or serving as targets for the  
CC development of human therapeutic agents that modulate tumour suppressor  
CC protein activity in cells and tissues that express the tumour suppressor  
CC protein. Polypeptides of the invention are used for identifying agents  
CC that modulate their activity. They are useful for raising antibodies or  
CC eliciting an immune response; as a reagent in assays designed to  
CC quantitatively determine levels of the protein (or its binding partner  
CC or ligand) in biological fluids; or as markers for tissues in which the  
CC corresponding protein is preferentially expressed. The invention is also  
CC used as vaccines. The present sequence is human tumour suppressor gene  
CC located on chromosome 13.  
XX  
SQ Sequence 163350 BP; 40430 A; 36356 C; 37002 G; 39020 T; 10542 other;  
Query Match 16.3%; Score 35.4; DB 24; Length 163350;  
Best Local Similarity 49.7%; Pred No. 4.6;  
Matches 90; Conservative 0; Mismatches 91; Indels 0; Gaps 0;  
QY 37 GTAAGTTCATTACACGGAAGAACTTTAAAGATAATATGAAAGGTATTACCTGTGTTG 96  
Db 163047 GCAACTGCTATTCATGCAATCTTCTATTAATTTGCAATTTACCTCCAATT 162988  
QY 97 CCACGGGAATGGTAATATGCCGAGTTTTTCACTGAATAGCAATCCAGCCATTCTTA 156  
Db 162987 AATATAGGAAGTGAAATAATTAGCATGAATTTTATATAGCAACATCAAGCTATTTTA 162928  
QY 157 TTCATATTTGACTGCGATCGCGTAATGTGGACTTTTATAGATAATGACCATGAAGATTTAAT 216  
Db 162927 TTGTCACGTGGATGTTTATTAAAGTATTTAGTTTCATTAACCTTTTGT 162868  
QY 217 T 217  
Db 162867 T 162867  
RESULT 14  
AAS45349/c  
ID AAS45349 standard; DNA; 17421 BP.  
XX  
AC AAS45349;  
DT 18-DEC-2001 (first entry)  
XX  
DE Chemically pretreated complementary DNA associated with cell cycle #27.  
XX  
KW Cell cycle; human; CpG dinucleotide; cytosine methylation; HIV; aging;  
KW human immunodeficiency virus; neurodegenerative disorder; solid tumour;  
KW graft-versus-host disease; glomerular disease; Lewy body disease; cancer;  
KW arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiarthritic;  
KW immunosuppressive; antitumour; cytostatic; antiarteriosclerotic; ds;  
KW PCR primer.  
XX  
OS Homo sapiens.  
XX  
PN WO200168911-A2.  
XX  
PD 20-SEP-2001.  
XX  
PF 15-MAR-2001; 2001WO-EP02945.  
XX  
PR 15-MAR-2000; 2000DE-1013847.  
PR 06-APR-2000; 2000DE-1019058.  
PR 07-APR-2000; 2000DE-1019173.  
PR 30-JUN-2000; 2000DE-1032529.  
PR 01-SEP-2000; 2000DE-1043826.  
XX  
PA (EPIG-) EPIGENOMICS AG.  
XX  
PI Olek A, Piepenbrock C, Berlin K;

XX WPI; 2001-602751/58.  
XX  
XX Designing primers and probes for analysing diseases associated with  
PT cytosine methylation state e.g. arthritis, cancer, aging,  
PT arteriosclerosis comprising fragments of chemically modified genes  
PT associated with cell cycle -  
XX  
XX Claim 1; SEQ ID No 54; 28pp; English.  
PS  
CC Sequences AAS45296-AAS45520 represent chemically pretreated genomic DNA  
CC molecules associated with the cell cycle and specific PCR primers of the  
CC invention. The sequences are useful for detecting the methylation state  
CC of all CpG dinucleotides in a sequence and therefore for analysing  
CC associated diseases. By analysing cytosine methylations in the pretreated  
CC DNA, genetic and/or epigenetic parameters for the diagnosis and therapy  
CC of existing diseases or the predisposition to specific diseases can be  
CC ascertained. The parameters may be compared to another set of genetic  
CC and/or epigenetic parameters, the differences serving as basis for  
CC diagnosis and/or prognosis events which are disadvantageous to patients.  
CC The sequences of the invention are useful for the diagnosis and therapy  
CC of HIV infection, neurodegenerative disorders, graft-versus-host disease,  
CC aging, glomerular disease, Lewy body disease, arthritis,  
CC arteriosclerosis, solid tumours and cancers.  
XX  
SQ Sequence 17421 BP; 4578 A; 328 C; 3973 G; 8502 T; 40 other;  
Query Match 16.0%; Score 34.8; DB 22; Length 17421;  
Best Local Similarity 65.4%; Pred No. 4.1;  
Matches 51; Conservative 0; Mismatches 27; Indels 0; Gaps 0;  
QY 16 AATACAGCGGATTAATACAGGTAAAGTTCATTACACGGAAGAACTTTAAAGATAATA 75  
Db 6042 AATATAAATAATATTATACCATATAATTTTATTTTACCATATAAATAAATAATAATA 5983  
QY 76 TGAACGGTATTACCTTGT 93  
Db 5982 TATAAATATATCTTTT 5965  
RESULT 15  
ABK28182/c  
ID ABK28182 standard; DNA; 17421 BP.  
XX  
AC ABK28182;  
DT 23-APR-2002 (first entry)  
XX  
DE DNA transcription associated complementary genomic DNA #28.  
XX  
KW DNA transcription associated gene; peptide nucleic acid; PNA-oligomer;  
KW PNA; cytosine methylation state; SNP; retroviral infection; gene; ds;  
KW single nucleotide polymorphism; adenosine deaminase deficiency; cancer;  
KW viral infection; Sezary syndrome; haematological disorder; tuberculosis;  
KW immunological disorder; Werner syndrome; developmental disorder;  
KW psoriasis; Rieger's syndrome; neurological disorder; erythropoiesis;  
KW neurodegenerative disorder; Waardenburg syndrome; Niemann-Pick disease;  
KW myelodysplastic syndrome; myocardial infarction; hypertension; arthritis;  
KW angiogenesis; congenital heart disease; HDR syndrome; gene therapy;  
KW polyglutamine disorder; solid tumour.  
XX  
OS Unidentified.  
XX  
PN WO200192565-A2.  
XX  
XX  
XX 06-DEC-2001.  
XX  
XX 06-APR-2001; 2001WO-EP03973.  
XX  
XX 06-APR-2000; 2000DE-1019058.  
XX 07-APR-2000; 2000DE-1019173.  
XX 30-JUN-2000; 2000DE-1032529.  
XX 01-SEP-2000; 2000DE-1043826.



XX (EPIC-) EPIGENOMICS AG.  
XX PA  
XX PI Olek A, Piepenbrock C, Berlin K;  
XX WPI; 2002-090046/12.  
XX DR  
XX PT New nucleic acids or oligomers, useful for diagnosing or treating  
XX diseases associated with DNA transcription, e.g. immunological  
XX disorders, Werner syndrome, psoriasis, myocardial infarction, solid  
XX tumours or cancer -  
XX  
PS Claim 1; SEQ ID No 56; 32pp; English.  
XX  
XX The invention relates to a nucleic acid, which comprises a segment of the  
XX chemically pretreated DNA of genes associated with DNA transcription from  
XX one of 346 sequences, and an oligomer, in particular an oligonucleotide  
XX or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical  
XX to the chemically pretreated DNA of genes associated with DNA  
XX transcription. The set of oligomer probes are useful for detecting the  
XX cytosine methylation state and/or single nucleotide polymorphisms (SNPs)  
XX in a chemically pretreated genomic DNA. The nucleic acids are useful for  
XX diagnosing or treating diseases associated with DNA transcription  
XX (particularly with the methylation status), e.g. adenosine deaminase  
XX deficiency, viral infection, retroviral infection, Sezary syndrome,  
XX haematological disorders, immunological disorders, Werner syndrome,  
XX tuberculosis, developmental disorders, psoriasis, Rieger's syndrome,  
XX neurological disorders, neurodegenerative disorders, Waardenburg  
XX syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial  
XX infarction, hypertension, angiogenesis, erythropoiesis, congenital heart  
XX disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumours  
XX or cancer. Sequences ABK28127-ABK28472 represent DNA transcription  
XX associated genomic DNA molecules of the invention.  
XX Note: The sequence data for this patent did not form part of the printed  
XX specification but was obtained in electronic format directly from the  
XX European Patent Office.  
SQ Sequence 17421 BP; 4578 A; 328 C; 3973 G; 8502 T; 40 other;  
Query Match 16.0%; Score 34.8; DB 24; Length 17421;  
Best Local Similarity 65.4%; Pred No. 4.1;  
Matches 51; Conservative 0; Mismatches 27; Indels 0; Gaps 0;  
QY 16 AATACAGCGGATTAATACAGGTAGTTCATTACACCGAAACCTTTAAAGATAATA 75  
DB 6042 AATATAAATATTAATACCATATATTTTATTACCATATAAAACATAAAATAATA 5983  
QY 76 TGAAGGTATTACCTTGT 93  
DB 5982 TATAAATTATATCTTTT 5965  
Search completed: November 15, 2003, 00:35:51  
Job time : 141.052 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 14, 2003, 23:58:39 / Search time 1447.63 Seconds  
(without alignments)  
3643.257 Million cell updates/sec

Title: US-09-928-457-99

Perfect score: 217

Sequence: 1 AATTGGACAGTATGAATAC.....ATGACGATGAAGATTAAATT 217

Scoring table:

IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rod:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	39.2	18.1	445	12	BM796121
C 2	39.2	18.1	682	14	CB851528
C 3	38.8	17.9	556	14	CA439981
C 4	38.6	17.8	600	28	AZ523328

RESULT 1  
BM796121/c  
LOCUS  
DEFINITION K-EST0078756 S22SNU16n1 Homo sapiens cDNA clone S22SNU16n1-53-C04  
5', mRNA sequence.  
ACCESSION BM796121  
VERSION BM796121.1 GI:19144353  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 445)  
AUTHORS Kim,N.S., Kahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.P., Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and Kim,Y.S.  
TITLE 21C Frontier Korean EST Project 2001  
JOURNAL Unpublished  
COMMENT Contact: Kim YS  
Genome Research Center  
Korea Research Institute of Bioscience & Biotechnology  
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea  
Tel: +82-42-860-4470  
Fax: +82-42-860-4409  
Email: yongsung@mail.kribb.re.kr  
Plate: 53 row: C column: 04  
High quality sequence stop: 445.

#### ALIGNMENTS

5	38.6	17.8	956	28	BH146874	BH146874	ENTQ15TR
C 6	38.2	17.6	380	9	AA294909	AA294909	EST100072
C 7	37.6	17.3	308	9	AA356543	AA356543	EST65107
C 8	37.6	17.3	312	14	T32582	EST51185	Hu
C 9	37.6	17.3	324	9	AA294908	AA294908	EST100071
C 10	37.6	17.3	340	14	RS9154	Y44905	xl
C 11	37.6	17.3	402	12	B1017065	B1017065	PM3-ET027
C 12	37.6	17.3	407	10	BF446491	7q88a03	x
C 13	37.6	17.3	435	12	BM755155	BM755155	K-EST0032
C 14	37.6	17.3	437	14	N57411	Y*82C07	xl
C 15	37.6	17.3	441	14	H48409	Y*32f06	xl
C 16	37.6	17.3	489	10	BF565645	UI-R-BT1-	
C 17	37.6	17.3	511	14	CB120144	CB120144	K-EST0167
C 18	37.6	17.3	515	10	AW959021	EST371091	
C 19	37.6	17.3	518	12	BM888435	BM888435	TM094 Hum
C 20	37.6	17.3	541	12	BM747615	BM747615	K-EST0022
C 21	37.6	17.3	596	12	BM744905	BM744905	K-EST0018
C 22	37.6	17.3	602	12	BM785453	BM785453	K-EST0063
C 23	37.6	17.3	620	12	BM743468	BM743468	K-EST0016
C 24	37.6	17.3	647	12	BM761014	BM761014	K-EST0041
C 25	37.6	17.3	675	12	BM706910	UI-E-C00-	
C 26	37.6	17.3	676	14	CB115165	CB115165	K-EST0159
C 27	37.6	17.3	682	9	AV682512	AV682512	AV682512
C 28	37.6	17.3	684	12	BM782216	BM782216	K-EST0059
C 29	37.6	17.3	754	10	BG576801	BG576801	602599137
C 30	37.6	17.3	775	14	CB963866	CB963866	AGENCOURT
C 31	37.6	17.3	782	10	BE545004	BE545004	601070073
C 32	37.6	17.3	819	10	BG168741	BG168741	602320021
C 33	37.6	17.3	830	13	BU933856	BU933856	AGENCOURT
C 34	37.6	17.3	832	12	B1260613	B1260613	602968101
C 35	37.6	17.3	848	10	BG171690	BG171690	602322213
C 36	37.6	17.3	864	13	BQ433011	BQ433011	AGENCOURT
C 37	37.6	17.3	873	13	BQ953645	BQ953645	AGENCOURT
C 38	37.6	17.3	880	13	BQ430231	BQ430231	AGENCOURT
C 39	37.6	17.3	883	13	BQ435954	BQ435954	AGENCOURT
C 40	37.6	17.3	884	13	BQ217267	BQ217267	AGENCOURT
C 41	37.6	17.3	899	13	BUI178416	BUI178416	AGENCOURT
C 42	37.6	17.3	920	13	BX393445	BX393445	EX393445
C 43	37.6	17.3	923	13	BX393444	BX393444	EX393444
C 44	37.6	17.3	925	10	BE782607	BE782607	601465726
C 45	37.6	17.3	925	10	BE910626	BE910626	601502609

FEATURES		Location/Qualifiers	
source		1..445	
		/organism="Homo sapiens"	
		/mol_type="mRNA"	
		/db_xref="taxon:9606"	
		/clone="S22SNU16n1-53-C04"	
		/sex="F"	
		/tissue_type="Ascites"	
		/cell_type="Lymphoblast-like"	
		/cell_line="SNU-16"	
		/lab_host="DH10B"	
		/clone_lib="S22SNU16n1"	
		/note="Organ: Stomach; Vector: pT73-Pac; Site 1: EcoRI; Site 2: NotI; The S22SNU16 library was contributed by the Soares laboratory and it was constructed as described by Bonaldo, M.F., Lennon, G. and Soares, M.B. (1996), Genome Research 6(9): 791-806. RNA was prepared from harvested cells of SNU-16 culture. SNU-16 cell was obtained from Korean Cell Line Bank (KCLB). SNU-16 was established from ascitic fluids of Korean patients by Park J.G. et al. (1990), Cancer Res 50: 2773-2780."	
BASE COUNT		108 a 75 c 90 g 172 t	
ORIGIN			
Query Match		18.1%; Score 39.2; DB 12; Length 445;	
Best Local Similarity		52.4%; Pred. No. 14;	
Matches		86; Conservative 0; Mismatches 78; Indels 0; Gaps 0;	
QY	1	AATCGGCACATGATACAGCGGATTATACAGGTAAGTTCATTACACGGAAAAAC	60
DB	236	AACTGAGGCATTAATACCTCTTTATATAATGCAAGTTGAATGCTAACAAAGCATAAAC	177
QY	61	CTTTAAAGAATAATAGAAAGTATTACCTTGTTGCCAACGGGAATGGTAATATGCC	120
DB	176	ACTCTGCAAAATTCACAGGGCACAGTGTTCATTACACAGAAAAGTCAAACCACT	117
QY	121	GAGTTTTCTCATGATACGGAATCCAGCCATTCTTCATTCATTT	164
DB	116	TGGTTTTTAATGAAATCCCTTCATCCACCTGTTCTCAAAAT	73
RESULT 2			
CB851528/c		682 bp mRNA linear EST 22-APR-2003	
LOCUS		UI-CF-EN1-aef-o-19-0-UI.s1 UI-CF-EN1 Homo sapiens cDNA clone	
DEFINITION		UI-CF-EN1-aef-o-19-0-UI 3', mRNA sequence.	
ACCESSION		CB851528	
VERSION		CB851528.1 GI:30046303	
KEYWORDS		EST.	
SOURCE		Homo sapiens (human)	
ORGANISM		Homo sapiens	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
AUTHORS		1 (bases 1 to 682)	
TITLE		Bonaldo, M.F., Lennon, G. and Soares, M.B.	
JOURNAL		Normalization and subtraction: two approaches to facilitate gene discovery	
MEDLINE		Genome Res. 6 (9), 791-806 (1996)	
PUBMED		97044477	
COMMENT		8889548	
		Contact: McCray, PB	
		McCray Lab	
		University of Iowa	
		2024 University of Iowa Med Labs, Iowa City, IA 52242, USA	
		Tel: 319 356 4866	
		Fax: 319 356 7171	
		Email: paul-mccray@uiowa.edu	
		Tissue Procurement: Dr. M. J. Welsh, University of Iowa	
		cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa	
		DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa	
		DNA Sequencing by: Dr. M. Bento Soares, University of Iowa	
		Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).	
Seq primer: M13 FORWARD			
POLYA=NO.			
FEATURES		Location/Qualifiers	
source		1..682	
		/organism="Homo sapiens"	
		/mol_type="mRNA"	
		/db_xref="taxon:9606"	
		/clone="UI-CF-EN1-aef-o-19-0-UI"	
		/tissue_type="Primary Lung Cystic Fibrosis Epithelial Cells"	
		/dev_stage="Adult"	
		/lab_host="DH10B (Life Technologies) (T1 phage resistant)"	
		/clone_lib="UI-CF-EN1"	
		/note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; UI-CF-EN1 is a normalized cDNA library containing the following tissue(s): Primary Lung Cystic Fibrosis Epithelial Cells. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CTGCTCAGGT.	
BASE COUNT		189 a 104 c 138 g 251 t	
ORIGIN			
Query Match		18.1%; Score 39.2; DB 14; Length 682;	
Best Local Similarity		52.4%; Pred. No. 15;	
Matches		86; Conservative 0; Mismatches 78; Indels 0; Gaps 0;	
QY	1	AATTCGGACAGTATGAATACAGCGGATTATACAGGTAAGTTCATTACACGGAAAAAC	60
DB	657	AACTGAGGCATTAATACCTCTTTATATAATGCAAGTTGAATGCTAACAAAGCATAAAC	598
QY	61	CTTTAAAGAATAATAGAAAGTATTACCTTGTTGCCAACGGGAATGGTAATATGCC	120
DB	597	ACTTCTGCAAAATTCACAGGGCACAGTGTTCATTACACAGAAAAGTCAAACCACT	538
QY	121	GAGTTTTCTCATGATACGGAATCCAGCCATTCTTCATTCATTT	164
DB	537	TGGTTTTTAATGAAATCCCTTCATCCACCTGTTCTCAAAAT	494
RESULT 3			
CA439981/c		556 bp mRNA linear EST 08-NOV-2002	
LOCUS		UI-H-D10-auf-1-07-0-UI.s1 NCI CGAP_D10 Homo sapiens cDNA clone	
DEFINITION		UI-H-D10-auf-1-07-0-UI 3', mRNA sequence.	
ACCESSION		CA439981	
VERSION		CA439981.1 GI:24804401	
KEYWORDS		EST.	
SOURCE		Homo sapiens (human)	
ORGANISM		Homo sapiens	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
AUTHORS		1 (bases 1 to 556)	
TITLE		NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.	
JOURNAL		National Cancer Institute, Cancer Genome Anatomy Project (CGAP),	
COMMENT		Tumor Gene Index	
		Unpublished	
		Contact: Robert Strausberg, Ph.D.	
		Email: cgapbs-remail.nih.gov	
		Tissue Procurement: Dr. Jose Mercuende	
		cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa	
		DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa	
		DNA Sequencing by: Dr. M. Bento Soares, University of Iowa	
		Clone Distribution: Clone distribution information can be obtained from Dr. M. Bento Soares, bento-soares@uiowa.edu	

The following repetitive elements were found in this cDNA  
 sequence: 526-556...>at\_richflow\_complexity  
 Seq primer: M13 FORWARD  
 POLYA=yes.

## FEATURES

## Location/Qualifiers

## source

```

1..556
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-H-D10-aux-1-07-0-UI"
/tissue_type="Lung Focal Fibrosis"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies)"
/clone_lib="NCI CGAP D10"
/notes="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; NCI-CGAP_D10 is a cDNA library containing the following tissue(s): A pool of lung Focal Fibrosis. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (GT)18 tail. The sequence tag for this library is ATACGGGTC.
TAG_LIB=UI-H-D10
TAG_TISSUE=lung with fibrosis
TAG_SEQ=ATACGGGTC"
130 a 106 c 117 g 203 t

```

## BASE COUNT

ORIGIN 130 a 106 c 117 g 203 t

```

Query Match 17.9%; Score 38.8; DB 14; Length 556;
Best Local Similarity 52.5%; Pred. No. 18;
Matches 85; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 1 AATCGGACAGTATGAATACAGCGGATTAATACAAAGTAAAGTTCAATACACGGAAAAAC 60
Db 177 AACTGGGCGATTAAATACCTCTTTATAATATCAAGTTGAAATGCTACAAAGCATAAAC 118

QY 61 CTTTAAAGATAATGAAGTATTAACCTGTTTGCCACGGGAATGGTAAATATGCC 120
Db 117 ACTTTGAAAAATTCACAGGCGACAGTTGTTCATCAACAGAAAAAGTCAAAAACCACT 58

QY 121 GAGTTTTCACATGATAGCAATCCAGCAATTCATTTCATA 162
Db 57 TGGTTTAAATGAAATCCITTCACATCCACTGTGTTCAA 16

```

## RESULT 4

```

A2525328
LOCUS 600 bp DNA linear GSS 07-MAY-2001
DEFINITION 242PbG08 pb MEN #21 Plasmodium berghei genomic 3', genomic survey sequence.
ACCESSION A2525328
VERSION A2525328.1 GI:13966072
KEYWORDS GSS.
SOURCE Plasmodium berghei
ORGANISM Plasmodium berghei
REFERENCE 1 (bases 1 to 600)
AUTHORS Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
TITLE Cartlon,J.M.-R. and Dame,J.B.
JOURNAL The Plasmodium vivax and P. berghei gene sequence tag projects
COMMENT Parasitol. Today (Regul. Ed.) 16 (10), 409 (2000)
Contact: Dame JB
Dept. of Pathobiology, College of Veterinary Medicine
University of Florida
2015 SW 23rd Avenue, Bldg 1017, Gainesville, FL 32611, USA
Tel: 352 392 4700
Fax: 352 392 9704
Email: damej@mail.vetmed.ufl.edu

```

Seq primer: M13(-20) forward  
 Class: shotgun.

## FEATURES

## Location/Qualifiers

## source

```

1..600
/organism="Plasmodium berghei"
/mol_type="genomic DNA"
/strain="ANKA clone 15cyl (clone of the ANKA 8417 clone)"
/db_xref="taxon:5821"
/dev_stage="asexual blood forms"
/lab_host="Mus musculus"
/clone_lib="PB MEN #21"
/notes="Vector: pBluescript SK(+), phagemid excised from lambda ZAP; Site 1: EcoRV; Site 2: EcoRV; Genomic DNA was prepared from asynchronous blood stage forms of the cloned ANKA isolate of P. berghei grown in laboratory Swiss white mice. The DNA was purified from contaminating host DNA by Hoechst Dye 33258-CsCl ultracentrifugation and precipitated. Purified DNA was digested with mung bean nuclease in the presence of 36-38% formamide at 50 C, as described (Vernick, K.D., Imberski, R.B., and McCutchan, T.F. 1988. Nucleic Acids Research 16:6883-6896). The ends of the digestion fragments were polished using T4 DNA polymerase, and the fragments size selected in the range 500-2000 bp. These were ligated into the EcoRV-cleaved and dephosphorylated pBluescript SK(+) vector. Recombinant plasmids were used to transform E. coli XL10-Gold host cells."
252 a 51 c 90 g 205 t 2 others

```

Query Match 17.8%; Score 38.6; DB 28; Length 600;  
 Best Local Similarity 49.0%; Pred. No. 21;  
 Matches 101; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

```

QY 12 TATGAATACAGCGGATTATACAGGTAAGTTCAATACACGGAAAAACCTTTAAAGAT 71
Db 366 TAAATAATGCCAAATAACATCAAGGGTTAGTAGACAAATAATAATCTATATCTATA 425

QY 72 AATATGAAGGTATTACCTTGTTCACACGGGAATGGTAAATATGCCGAGTTTTCAC 131
Db 426 AAGATAAGTGGATGAAATTTAATCAAAATGTTTTCAGTATATGCACATAAATTTCCAA 485

QY 132 TGAATAGCGAATCCAGCCATTTCTTATTCATATTTGACTGGATGGCTGAATGTGACTTTA 191
Db 486 GGCATGCGAATCAAGGATGGCAAAATATAAANGCAACTTAATTTGTTGATGATCTAA 545

QY 192 TAGATAATGACGATGAGATTTAATT 217
Db 546 TTGATGATGATATACAAATTTTGAAT 571

```

## RESULT 5

```

BH146874
LOCUS 956 bp DNA linear GSS 27-AUG-2001
DEFINITION ENTQ15TR Entamoeba histolytica sheared DNA Entamoeba histolytica genomic, genomic survey sequence.
ACCESSION BH146874
VERSION BH146874.1 GI:15302923
KEYWORDS GSS.
SOURCE Entamoeba histolytica
ORGANISM Entamoeba histolytica
REFERENCE 1 (bases 1 to 956)
AUTHORS Eukaryota; Entamoebidae; Entamoeba.
TITLE Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.
JOURNAL Determination of clone end sequences from Entamoeba histolytica
COMMENT HM1:IMSS sheared DNA library (2001)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543

```

Email: bjloftus@tigr.org  
 Clones are derived from the Entamoeba histolytica HM1:IMSS sheared  
 DNA library  
 Seq primer: M13-Reverse  
 Class: shotgun  
 High quality sequence start: 47  
 High quality sequence stop: 458  
 Location/Qualifiers

## FEATURES

1..956  
 /organism="Entamoeba histolytica"

/mol\_type="genomic DNA"

/strain="HM1:IMSS"

/db\_xref="taxon:5759"

/clone\_lib="Entamoeba histolytica Sheared DNA"

/note="Vector: pBOS1; Site 1: Bst I; Constructed at The

Institute for Genomic Research (TIGR), Rockville, MD.

Genomic DNA isolated from broth cultures of E. histolytica

using a method described by Clark and Diamond (Clark, a

C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a

method for isolate identification. Exp. Parasitol.

77:450.). The DNA was mechanically sheared to give a

tight size distribution (~2 kb). The v + i method used for

the library construction is described in detail in Smith,

H.O. and Venter, J.C. (Making small insert libraries for

whole genome shotgun sequencing projects. In Genome

Sequencing: A Practical Approach, eds. M. Vaudin and B.

Barell, Oxford University Press, 1995)."

BASE COUNT 311 a 158 c 226 g 261 t

## ORIGIN

Query Match 17.8%; Score 38.6; DB 28; Length 956;  
 Best Local Similarity 56.8%; Pred. No. 22;  
 Matches 71; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 25 GATTATACAGCTAAGTTTCATTACACGGAAAAACCTTTAAAGATAATATGAAGCTA 84

DB 575 GATTGAATAGTTTAAGAAACCTCAAAGACGCTACCTGACGGGATAATATGGAA 634

QY 85 TTACCTGTTGCCAACGGGAATGTAATATGCCGAGTGTTCCTGCTGATGACGATC 144

DB 635 TTATTGTTTTTGAACAAGGAATTCCTCGAATTTCCAAAGTCTTTTACTCCAATGGAATA 694

QY 145 CAGCC 149

DB 695 CGGCC 699

## RESULT 6

AA294909/c  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE

AA294909 Pancreas tumor I Homo sapiens cDNA 5' end, mRNA sequence.  
 AA294909.1 GI:1947264

EST.  
 Homo sapiens (human)  
 Homo sapiens

ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  
 AUTHORS

1 (bases 1 to 380)  
 Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fullmer, R.A., Bult

, C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D., White

, C., Sutton, G., Blake, J.A., Brandon, R.C., Man-Wai, C., Clayton, R.A.,

Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fine, L.D., Fitzgerald

, L.M., Fitzhugh, W.M., Fitchman, J.L., Geoghagen, N.S., Glodek, A.,

Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S., Jr., Kelley, J.M.,

Kelly, J.C., Liu, L.-I., Marmaro, S.M., Merrick, J.M.,

Moreno-Palancas, R.F., McDonald, L.A., Nguyen, D.T., Pallgrino, S.M.,

Phillips, C.A., Ryder, S.E., Scott, J.L., Saudek, D.M., Shirley, R.,

Wei, Y.F., Wing, J., Xu, C., Yu, G.L., Ruben, S.M., Dillion, P.J., Pannon

, M.R., Rosen, C.A., Haseltine, W.A., Fields, C., Fraser, C.M. and

Venter, J.C.  
 Initial assessment of human gene diversity and expression patterns

based upon 83 million nucleotides of cDNA sequence

Nature 377 (6547 Suppl), 3-174 (1995)

Other ESTs: THC181566  
 96026280  
 7566098

Contact: Kerlavage, AR  
 Bioinformatics  
 The Institute for Genomic Research

9712 Medical Center Drive, Rockville, MD 20850 USA

Tel: 3018699056  
 Fax: 3018699423

Email: arkerlav@tigr.org  
 For clone availability, additional sequence and expression

information related to this EST, please check the TIGR Human Gene

Index (<http://www.tigr.org/tdb/hgi/hgi.html>)

Seq primer: M13 Reverse  
 Location/Qualifiers

1..380  
 /organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="ATCC (inhost):190411"

/db\_xref="taxon:9606"

/dev\_stage="adult"

/clone\_lib="pancreas tumor I"

/note="Organ: pancreas; Vector: pBluescript SK-; Site\_1:

ECORI; Site\_2: XhoI"

BASE COUNT 99 a 48 c 79 g 152 t 2 others

ORIGIN

Query Match 17.6%; Score 38.2; DB 9; Length 380;  
 Best Local Similarity 51.8%; Pred. No. 25;  
 Matches 85; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 1 AATTCGACAGTATGATACAGCGGATTAATCAAGTGAAGTTCATTACACGGAATAAC 60

DB 295 AACTGAGGATTAATACCTCTTTATATAATCAAGTTGAAATGCTNACAAAGCATAAC 236

QY 61 CTTTAAAGATAATATGAAGGTATACCTTGTTCGCAACGGATGTTAAATATGCC 120

DB 235 ACTTCTGCAAAAATTCACAGGCACAGTTGTTCATTCAACAGAAAAGTCAAAACCACT 176

QY 121 GAGTTTTTCAGTGAATAGCAATCCAGCCATTTCTTATTCATATT 164

DB 175 TGGTTTTTAATGAANAATCCTTCACATCCACCTGTTCAAAAT 132

RESULT 7  
 AA356543/c  
 LOCUS

DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE

AA356543 Jurkat T-cells VI Homo sapiens cDNA 5' end, mRNA sequence.  
 AA356543.1 GI:2008862

EST.  
 Homo sapiens (human)  
 Homo sapiens

ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  
 AUTHORS

1 (bases 1 to 308)  
 Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fullmer, R.A., Bult

, C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D., White

, C., Sutton, G., Blake, J.A., Brandon, R.C., Man-Wai, C., Clayton, R.A.,

Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fine, L.D., Fitzgerald

, L.M., Fitzhugh, W.M., Fitchman, J.L., Geoghagen, N.S., Glodek, A.,

Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S., Jr., Kelley, J.M.,

Kelly, J.C., Liu, L.-I., Marmaro, S.M., Merrick, J.M.,

Moreno-Palancas, R.F., McDonald, L.A., Nguyen, D.T., Pallgrino, S.M.,

Phillips, C.A., Ryder, S.E., Scott, J.L., Saudek, D.M., Shirley, R.,

Bednarik, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J.,

Dimke, D., Feng, D.-F., Ferrie, A., Fischer, C., Hastings, G.A., He, W.W.

Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K., Kozak, D.L.,

Kunsch, C., Hungjun, J., Li, H., Meissner, P.S., Olsen, H., Raymond, L.,

Bednarik, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J.,

Dimke, D., Feng, D.-F., Ferrie, A., Fischer, C., Hastings, G.A., He, W.W., Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K., Kozak, D.L., Kunsch, C., Hung, J., Li, H., Meissner, P.S., Olsen, H., Raymond, L., Wei, Y.F., Wing, J., Xu, C., Yu, G.L., Ruben, S.M., Dillon, P.J., Fannon, M.R., Rosen, C.A., Haseltine, W.A., Fields, C., Fraser, C.M. and Venter, J.C.

Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence  
Nature 377 (6547 Suppl), 3-174 (1995)

TITLE  
JOURNAL  
MEDLINE  
PUBMED  
COMMENT  
Other ESTs: THC181566  
Contact: Kerlavage, AR  
Bioinformatics  
The Institute for Genomic Research  
9712 Medical Center Drive, Rockville, MD 20850 USA  
Tel: 3018699056  
Fax: 3018699423  
Email: arkerlav@tigr.org  
For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (<http://www.tigr.org/tdb/hgi/hgi.html>)  
Seq primer: M13 Reverse.  
Location/Qualifiers  
1. 308  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="Arcc (inhost):158684"  
/cell\_type="T-lymphocyte"  
/clone\_lib="Jurkat T-cells VI"  
/note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2: XhoI"  
XhoI  
90 a 32 c 57 g 129 t

FEATURES  
source  
1. 308  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="Arcc (inhost):158684"  
/cell\_type="T-lymphocyte"  
/clone\_lib="Jurkat T-cells VI"  
/note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2: XhoI"  
XhoI  
90 a 32 c 57 g 129 t

BASE COUNT  
ORIGIN  
Query Match 17.3%; Score 37.6; DB 9; Length 308;  
Best Local Similarity 51.8%; Pred No. 34;  
Matches 85; Conservative 0; Mismatches 79; Indels 0; Gaps 0;  
QY 1 AATTCGGACAGTAAATACAGCGGATTAAATACAGGTAAGTTCAATCAACGGAAAC 60  
Db 300 AACTCGGACATTATACCTCTTTATATATGCAAGTTGAAATGCTAACAAGCATTAAC 241  
QY 61 CTTTAAAGAAATATGAAGGTATTACCTTGTTCGCAAGGAAATGTAATATGCCC 120  
Db 240 ACTTCGCAAAATTCACAGGACAGTTGTTCAATCAACAGAAAGTCAAAACCACT 181  
QY 121 GAGTTTTCACTGAATAGCAATCCAGCCATTCTATTCAAT 164  
Db 180 TGGTTTTAAATGAAATCCCTTCACATCCACTGTGTTCAAAAT 137

RESULT 8  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
T32582  
EST511485 Human Gall bladder Homo sapiens cDNA 5' end similar to None. mRNA sequence.  
T32582  
T32582.1 GI:614680  
EST.  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 312)

REFERENCE  
AUTHORS  
Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A., Bult, C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D., White, O., Sutton, G., Blake, J.A., Brandon, R.C., Chiu, M.-W., Clayton, R.A., Cline, R.T., Cotton, M.D., Earle-Hughes, J., Fine, L.D., FitzGerald, L.M., Fitchugh, W.M., Fritchman, J.L., Geoghagen, N.S.M., Glodek, A., Gnahn, C.L., Hanna, M.C., Hedblom, E., Hinkley, P.S., Kelley, J.M., Klimek, K.M., Kelley, J.C., Liu, L.-I., Marmaros, S.M., Merrick, J.M.,

Moreno-Palauques, R.F., McDonald, L.A., Nguyen, D.T., Pellegrino, S.M., Phillips, C.A., Ryder, S.E., Scott, J.L., Saudek, D.M., Shirley, R., Small, K.V., Sprigge, T.A., Uterback, T.R., Weidman, J.F., Li, Y., Bednarik, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.-J., Dimke, D., Feng, P., Ferrie, A., Fischer, C., Hastings, G.A., He, W.-W., Hu, J.-S., Greene, J.M., Gruber, J., Hudson, P., Kim, A., Kozak, D.L., Kunsch, C., Li, H., Li, H., Meissner, P.S., Olsen, H., Raymond, L., Wei, Y.-F., Wing, J., Xu, C., Yu, G.-L., Ruben, S.M., Dillon, P.J., Fannon, M.R., Rosen, C.A., Haseltine, W.A., Fields, C., Fraser, C.M. and Venter, J.C.

Initial Assessment of Human Gene Diversity and Expression Patterns Based Upon 83 Million Basepairs of cDNA Sequence  
Nature 377, 3-174 (1995)

TITLE  
JOURNAL  
MEDLINE  
PUBMED  
COMMENT  
Other ESTs: THC9475  
Contact: Venter, JC  
The Institute for Genomic Research  
932 Clopper Rd, Gaithersburg, MD 20878  
Tel: 3018699056  
Fax: 3018699423  
Email: tdbinfo@db.tigr.org  
For clone availability, additional sequence and expression information related to this EST, please contact the TIGR Database (tdbinfo@db.tigr.org)  
Seq primer: M13 Reverse.  
Location/Qualifiers  
1. 312  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="Arcc (inhost):103165"  
/db\_xref="taxon:9606"  
/clone\_lib="Human Gall bladder"  
/note="Organ: gallbladder"  
80 a 36 c 64 g 132 t

FEATURES  
source  
1. 312  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="Arcc (inhost):103165"  
/db\_xref="taxon:9606"  
/clone\_lib="Human Gall bladder"  
/note="Organ: gallbladder"  
80 a 36 c 64 g 132 t

BASE COUNT  
ORIGIN  
Query Match 17.3%; Score 37.6; DB 14; Length 312;  
Best Local Similarity 51.8%; Pred. No. 34;  
Matches 85; Conservative 0; Mismatches 79; Indels 0; Gaps 0;  
QY 1 AATTCGGACAGTAAATACAGCGGATTAAATACAGGTAAGTTCAATCAACGGAAAC 60  
Db 246 AACTCGGACATTATACCTCTTTATATATGCAAGTTGAAATGCTAACAAGCATTAAC 187  
QY 61 CTTTAAAGAAATATGAAGGTATTACCTTGTTCGCAAGGAAATGTAATATGCCC 120  
Db 186 ACCTTCGCAAAATTCACAGGACAGTTGTTCAATCAACAGAAAGTCAAAACCACT 127  
QY 121 GAGTTTTCACTGAATAGCAATCCAGCCATTCTATTCAAT 164  
Db 126 TGGTTTTAAATGAAATCCCTTCACATCCACTGTGTTCAAAAT 83

RESULT 9  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
AA294908  
EST100071 Pancreas tumor I Homo sapiens cDNA 5' end, mRNA sequence.  
AA294908  
AA294908.1 GI:1947263  
EST.  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 324)

REFERENCE  
AUTHORS  
Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A., Bult, C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D., White, O., Sutton, G., Blake, J.A., Brandon, R.C., Man-Wai, C., Clayton, R.A., Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fine, L.D., FitzGerald, L.M., Fitchugh, W.M., Fritchman, J.L., Geoghagen, N.S., Glodek, A., Gnahn, C.L., Hanna, M.C., Hedblom, E., Hinkley, P.S., Kelley, J.M., Kelley, J.C., Liu, L.-I., Marmaros, S.M., Merrick, J.M.,

Moreno-Palancas, R.F., McDonald, L.A., Nguyen, D.T., Pelligrino, S.M., Phillips, C.A., Ryder, S.E., Scott, J.L., Saudek, D.M., Shirley, R., Small, K.V., Spriggs, T.A., Utterback, T.R., Weidman, J.F., Li, Y., Bednarik, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, B.J., Dinkel, D., Feng, D.-F., Fertie, A., Fischer, C., Hastings, G.A., He, W.W., Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K., Kozak, D.L., Kunsch, C., Hungjun, J., Li, H., Meisner, P.S., Olsen, H., Raymond, L., Wei, Y.F., Wing, J., Xu, C., Yu, G.L., Ruben, S.M., Dillon, P.J., Fannon, M.R., Rosen, C.A., Haseltine, W.A., Fields, C., Fraser, C.M. and Venter, J.C.

Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence  
Nature 377 (6547 Suppl), 3-174 (1995)  
96026280  
7586098  
Other ESTs: THC181566  
PUBMED  
Contact: Kerlavage, AR  
Bioinformatics  
The Institute for Genomic Research  
9712 Medical Center Drive, Rockville, MD 20850 USA  
Tel: 301899056  
Fax: 301899423  
Email: arkerlav@tigr.org  
For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (<http://www.tigr.org/tdb/hgi/hgi.html>)  
Seq primer: M13 Reverse.

FEATURES  
source  
1..324  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="ATCC (inhost):190410"  
/db\_xref="taxon:9606"  
/dev\_stage="adult"  
/clone\_lib="Pancreas tumor 1"  
/note="Organ: pancreas; Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2: XhoI"  
BASE COUNT 90 a 35 c 64 g 133 t 2 others  
ORIGIN  
Query Match 17.3%; Score 37.6; DB 9; Length 324;  
Best Local Similarity 51.8%; Pred. No. 34;  
Matches 85; Conservative 0; Mismatches 79; Indels 0; Gaps 0;  
QY 1 AATTCGACAGTATGAATACACGGGATTAATACAGGTAAAGTTCAATACACGGGAAAAAC 60  
DB 295 AACTGAGCAATTAATACCTTTTATATATGCAAGTTGAATCTCAACAAAGCATAAAC 236  
QY 61 CTTTAAAGAAATATGAAGTATTACCTTTGTTGCCAAGGGAATGGTAATATGCC 120  
DB 235 ACTTCTGCAAAATTCACAAAGGCACAGTTGTTCAATCAACAGAAAAGTCAAAACCACT 176  
QY 121 GAGTTTTCACATGAATACGAAATCCAGCAATTCATTCATATT 164  
DB 175 TCGTTTAAATGAATCCCTTCATCCACCTGTGTTCAAAAT 132

RESULT 10  
R69154/c  
LOCUS  
DEFINITION  
IMAGE:142136 3', mRNA sequence.  
ACCESSION  
R69154  
VERSION  
R69154.1 GI:842671  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 340)  
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaaskis, E., Waterston

R., Williamson, A., Wohldmann, P. and Wilson, R.  
The WashU-Werck EST Project  
Unpublished  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.wustl.edu  
Insert Size: 966  
High quality sequence stops: 245  
Source: IMAGE Consortium, LLNL  
This clone is available royalty-free through LLNL; contact the IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Insert Length: 966 Std Error: 0.00  
Seq primer: Promega -21m13  
High quality sequence stop: 245.  
Location/Qualifiers  
1..340  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="GDB:550897"  
/db\_xref="taxon:9606"  
/clone="IMAGE:142136"  
/sex="Female"  
/dev\_stage="placenta obtained at birth (full term)"  
/lab\_host="DH10B (ampicillin resistant)"  
/clone\_lib="Soares placenta Nb2Hp"  
/note="Organ: placenta; Vector: p7T3D (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', RACTGGAAGAAATCGCGCGAGAAATTTTTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p7T3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 85 a 47 c 74 g 132 t 2 others  
ORIGIN  
Query Match 17.3%; Score 37.6; DB 14; Length 340;  
Best Local Similarity 51.8%; Pred. No. 35;  
Matches 85; Conservative 0; Mismatches 79; Indels 0; Gaps 0;  
QY 1 AATTCGACAGTATGAATACACGGGATTAATACAGGTAAAGTTCAATACACGGGAAAAAC 60  
DB 215 AACTGAGCAATTAATACCTTTTATATATGCAAGTTGAATCTCAACAAAGCATAAAC 156  
QY 61 CTTTAAAGAAATATGAAGTATTACCTTTGTTGCCAAGGGAATGGTAATATGCC 120  
DB 155 ACTTCTGCAAAATTCACAAAGGCACAGTTGTTCAATCAACAGAAAAGTCAAAACCACT 96  
QY 121 GAGTTTTCACATGAATAGCAATCCAGCAATTCATTCATATT 164  
DB 95 TCGTTTAAATGAATCCCTTCATCCACCTGTGTTCAAAAT 52

RESULT 11  
BI017065  
LOCUS  
DEFINITION  
PM3-ET0272-300301-002-b06 ET0272 Homo sapiens cDNA, mRNA sequence.  
ACCESSION  
BI017065  
VERSION  
BI017065.1 GI:14421136  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 402)  
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare



TITLE  
JOURNAL  
MEDLINE  
PUBMED  
COMMENT

M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.  
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
200202663  
10737800  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?l1=PM3&t2=PM3-ET0272-300301-002-b06&t3=2001-03-30&t4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 5  
High quality sequence stop: 402.

## FEATURES

Location/Qualifiers

source

1..402

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/dev\_stage="Adult"

/clone\_lib="ET0272"

/note="Organ: lung; tumor; Vector: puc18; Site: 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196.716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

163 a 79 c 58 g 101 t 1 others

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

85; Conservative

0; Mismatches

79; Indels

0; Gaps

0;

QY

1 AATTGGCAGATGAAATACAGCGGATTATACAGGTAAGTTCAATTCATTAACCGGAAAC 60

DB

120 AACTGAGCATTATACCTCTTTATATATGCAAGTTGAAATGCTAACAAAGCATAAAC 179

QY

61 CTTTAAAGTAATATGAAAGTATACCTTGTTCGCAAGGTAATGTTAAATATGCC 120

DB

180 ACTTCTGCAAAATTCACAGGCGACAGTTGTTCAATTCACAGAAAGTCAAAACCACT 239

QY

121 GAGTTTTCTACTGATAGCGAATCCAGCCATTTCTTATTCATATT 164

DB

240 TGGTTTTTAATGAAATCCCTTCACATCCACCTGTGTTCAAAAT 283

RESULT 12

BF446491/c

LOCUS

DEFINITION

BF446491

7q88a03.x1 NCI\_CGAP\_Lu24 Homo sapiens cDNA clone IMAGE:3705125 3', mRNA sequence.

ACCESSION

BF446491

VERSION

BF446491.1 GI:11511629

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 407)

REFERENCE

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

AUTHORS

National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL

Unpublished

## COMMENT

Contact: Robert Strausberg, Ph.D.  
Email: cgabps@mail.nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov  
Seq primer: -40UP from Gibco  
High quality sequence stop: 405.  
Location/Qualifiers

## FEATURES

source

1..407

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:3705125"

/tissue\_type="carcinoid"

/lab\_host="DH10B"

/clone\_lib="NCI\_CGAP\_Lu24"

/note="Organ: lung; Vector: p7T3D-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI CGAP Lu5 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clonoids 1414920-1417991 and 1520904-1522439). Subtraction by Bento Soares and M. Fatima Bonaldo."

120 a 46 c 77 g 164 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

85; Conservative

0; Mismatches

79; Indels

0; Gaps

0;

QY

1 AATTGGCAGATGAAATACAGCGGATTATACAGGTAAGTTCAATTCATTAACCGGAAAC 60

DB

392 AACTGAGCATTATACCTCTTTATATATGCAAGTTGAAATGCTAACAAAGCATAAAC 333

QY

61 CTTTAAAGTAATATGAAAGTATACCTTGTTCGCAAGGTAATGTTAAATATGCC 120

DB

332 ACTTCTGCAAAATTCACAGGCGACAGTTGTTCAATTCACAGAAAGTCAAAACCACT 273

QY

121 GAGTTTTCTACTGATAGCGAATCCAGCCATTTCTTATTCATATT 164

DB

272 TGGTTTTTAATGAAATCCCTTCACATCCACCTGTGTTCAAAAT 229

RESULT 13

BM755155/c

LOCUS

DEFINITION

BM755155

435 bp mRNA linear EST 04-MAR-2002

K-EST0032842 S1LSNUL Homo sapiens cDNA clone S1LSNUL-7-C07 5', mRNA sequence.

ACCESSION

BM755155

VERSION

BM755155.1 GI:19084773

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 435)

REFERENCE

Kim, N.S., Hahn, Y., Oh, J.H., Lee, J.Y., Ahn, H.Y., Chu, M.Y., Kim, M.R., Oh, K.J., Cheong, J.E., Sohn, H.Y., Kim, J.M., Park, H.S., Kim, S. and Kim, Y.S.

21C Frontier Korean EST Project 2001

Unpublished

Contact: Kim YS

Genome Research Center

Korea Research Institute of Bioscience &amp; Biotechnology

52 Eeun-dong Yuseong-gu, Daejeon 305-333, South Korea

Tel: +82-42-860-4470



Search completed: November 15, 2003, 08:02:54  
Job time : 1451.63 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 15, 2003, 00:06:39 ; Search time 33.7995 Seconds  
(without alignments)  
2833.774 Million cell updates/sec

Title: US-09-928-457-99

Perfect score: 217

Sequence: 1 AATTCGGACAGTATGATAC.....ATGACGATGACATTAATT 217

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	37.6	17.3	3826	4 US-09-220-132-51	Sequence 51, Appl
C 2	34.2	15.8	2047	3 US-08-836-261A-1	Sequence 1, Appl
C 3	32.4	14.9	1830121	4 US-09-557-884-1	Sequence 1, Appl
C 4	32.4	14.9	1830121	4 US-09-643-990A-1	Sequence 1, Appl
C 5	32	14.7	1684976	4 US-08-916-421B-1	Sequence 1, Appl
C 6	31.6	14.6	1478	1 US-08-090-523-27	Sequence 27, Appl
C 7	31.6	14.6	1478	1 US-08-398-627-27	Sequence 27, Appl
C 8	31.6	14.6	1478	1 US-08-406-857-1	Sequence 1, Appl
C 9	31.6	14.6	1478	1 US-08-596-024-3	Sequence 3, Appl
C 10	31.6	14.6	1478	1 US-09-020-818-3	Sequence 3, Appl
C 11	31.6	14.6	1478	3 US-08-907-740-3	Sequence 3, Appl
C 12	31.6	14.6	1478	4 US-09-797-467-3	Sequence 3, Appl
C 13	31.6	14.6	1478	4 US-08-399-023-27	Sequence 27, Appl
C 14	31.6	14.6	1478	5 PCT-US94-07072-1	Sequence 1, Appl
C 15	31.2	14.4	3030	4 US-09-328-352-4115	Sequence 4115, Ap
C 16	30.6	14.1	1184	1 US-08-362-511A-1	Sequence 4, Appl
C 17	30.6	14.1	1239	1 US-08-362-511A-1	Sequence 1, Appl
C 18	30.6	14.1	1684976	4 US-08-916-421B-1	Sequence 1, Appl
C 19	30.2	13.9	7418	4 US-09-620-312D-365	Sequence 365, App
C 20	30	13.8	2142	4 US-09-107-532A-905	Sequence 905, App
C 21	30	13.8	148567	4 US-09-801-876B-3	Sequence 3, Appl
C 22	29.6	13.6	1172	4 US-08-969-046-5	Sequence 5, Appl
C 23	29.6	13.6	2255	3 US-08-714-918-105	Sequence 105, App
C 24	29.6	13.6	2255	3 US-09-265-315-105	Sequence 105, App
C 25	29.6	13.6	2255	3 US-09-265-315-105	Sequence 105, App
C 26	29.6	13.6	2255	3 US-09-266-417-105	Sequence 105, App
C 27	29.4	13.5	3600	3 US-08-894-731-1	Sequence 1, Appl

C 28	29.2	13.5	419	3 US-09-030-607-201	Sequence 201, App
C 29	29.2	13.5	419	4 US-09-439-313-201	Sequence 201, App
C 30	29.2	13.5	419	4 US-09-352-616A-201	Sequence 201, App
C 31	29.2	13.5	419	4 US-09-232-149A-201	Sequence 201, App
C 32	29.2	13.5	1230025	4 US-09-198-452A-1	Sequence 1, Appl
C 33	29.2	13.3	597	4 US-09-712-016-73	Sequence 73, Appl
C 34	28.6	13.2	1977	2 US-08-825-558-3	Sequence 3, Appl
C 35	28.6	13.2	1977	4 US-09-312-611-3	Sequence 3, Appl
C 36	28.6	13.2	2369	1 US-07-797-556-1	Sequence 1, Appl
C 37	28.6	13.2	2369	1 US-08-308-881-1	Sequence 1, Appl
C 38	28.6	13.2	2369	2 US-09-058-263-1	Sequence 1, Appl
C 39	28.6	13.2	2369	2 US-09-059-099-1	Sequence 1, Appl
C 40	28.6	13.2	2369	3 US-09-058-264-1	Sequence 1, Appl
C 41	28.6	13.2	2369	4 US-09-455-962-1	Sequence 1, Appl
C 42	28.6	13.2	2369	5 PCT-US95-06530-1	Sequence 1, Appl
C 43	28.6	13.2	2754	2 US-08-825-558-5	Sequence 5, Appl
C 44	28.6	13.2	2754	4 US-09-312-611-5	Sequence 5, Appl
C 45	28.6	13.2	3085	3 US-08-795-473B-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1  
US-09-220-132-51/c  
; Sequence 51, Application US/09220132  
; Patent No. 6506607  
; GENERAL INFORMATION:  
; APPLICANT: Shyjan, Andrew W.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE IDENTIFICATION AND ASSESSMENT  
; FILE REFERENCE: 07334-074001  
; CURRENT APPLICATION NUMBER: US/09/220,132  
; CURRENT FILING DATE: 1998-12-23  
; PRIOR APPLICATION NUMBER: US 60/079,303  
; PRIOR FILING DATE: 1998-03-25  
; PRIOR APPLICATION NUMBER: US 60/068,821  
; PRIOR FILING DATE: 1997-12-24  
; NUMBER OF SEQ ID NOS: 191  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 51  
; LENGTH: 3826  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; US-09-220-132-51

Query Match	17.3%;	Score 37.6;	DB 4;	Length 3826;
Best Local Similarity	51.8%;	Pred. No. 0.041;		
Matches	85;	Conservative	0;	Mismatches 79;
Indels	0;	Gaps	0;	
Qy	1	AATCGGACAGTATGATACAGCGGATTATACAAAGTAAAGTTTCATTACAAACGGAAC	60	
Db	3162	AAC TGAGGCATTAACTCTTTTATATAATGCAAGTTGAAATGCTTACAAAGCATAAAC	3103	
Qy	61	CTTTAAAGATATATGAAGGTATTACCTGTTGCCACGGGATGTAATATGCCC	120	
Db	3102	ACTTCGCAAAATTCACAAAGCAGTTGTTTCATTTCACAGAAAGTCAAAACCACT	3043	
Qy	121	GAGTTTTTCACTGAATAGCGAATCCAGCCATTTCATTTCATTT	164	
Db	3042	TGTTTTTAAATGAAATCTTCACATCCACCTGTGTTCAAAAT	2999	

RESULT 2  
US-08-836-261A-1  
; Sequence 1, Application US/08836261A  
; Patent No. 6221582  
; GENERAL INFORMATION:  
; APPLICANT: GIESENDORF, BELINDA  
; APPLICANT: CUINT, WILHELMUS  
; APPLICANT: VAN DOORN, LEENDERT-JAN  
; TITLE OF INVENTION: NEW POLYNUCLEIC ACID SEQUENCES FOR USE IN THE  
; DETECTION AND DIFFERENTIATION OF PROKARYOTIC ORGANISMS

```
;
;
; NUMBER OF SEQUENCES: 96
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARNOLD, WHITE & DURKEE
; STREET: P.O. BOX 4433
; CITY: HOUSTON
; STATE: TEXAS
; COUNTRY: USA
; ZIP: 77210-4433
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word 6.0 / ASCII text output
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/836.261A
; FILING DATE: 25 Apr 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP95/04264
; FILING DATE: 30 Oct 1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 94870171.9
; FILING DATE: 28 Oct 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: KAMMERER, PATRICIA A.
; REGISTRATION NUMBER: 29,775
; REFERENCE/DOCKET NUMBER: INNS:005
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2047 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 864..2016
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 864..2013
;
; US-08-836-261A-1
;
; Query Match 15.8%; Score 34.2; DB 3; Length 2047;
; Best Local Similarity 49.2%; Pred. No. 0.35;
; Matches 90; Conservative 0; Mismatches 93; Indels 0; Gaps 0;
;
; QY 34 AAGGTAAGTTCATTACACAGGAAACCTTTAAAGAAATAATATGAAAGGTATTACCTTGT 93
; Db 824 AATTTTATATAATTTTAAATAATAATTTTCTAGGAAAAAATGCAAAGCATCATCTTA 883
;
; QY 94 TTGCCAAGGGAATGGTAAATATGCCGAGTTTTCAGTGAATAGCGAATCCAGCCATTT 153
; Db 884 TAGGCAAGCCAAATGTTGGAAATCAAGTCCTTTTATAGAAATGCGAAGCAAGAAATAG 943
;
; QY 154 CTATTTCATATTGACTGGATGCTGAATGTCGACTTTTATAGATAATGACGATGAAGATTT 213
; Db 944 CTATTACAGTCATATTTCAGGTACACTAGATACAGAAATAAAGCCAAATTCATATTC 1003
;
; QY 214 AAT 216
; Db 1004 ATT 1006
;
;
; RESULT 3
; US-09-557-884-1
; Sequence 1, Application US/09557884
; Patent No. 6506581
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: The Nucleotide sequence of
; the Haemophilus influenzae Rd Genome, Fragments
; Thereof, and Uses Thereof
;
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fleischmann et al.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: MD
; COUNTRY: USA
; ZIP: 20850
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3 1/2 inch diskette
; COMPUTER: Dell Pentium
; OPERATING SYSTEM: MS DOS v6.22
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/557,884
; FILING DATE: 25-Apr-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/476,102
; FILING DATE: JUN-3-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB186P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-309-8504
; TELEFAX: 301-309-8439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1830121 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
;
; US-09-557-884-1
;
; Query Match 14.9%; Score 32.4; DB 4; Length 1830121;
; Best Local Similarity 48.9%; Pred. No. 8.2;
; Matches 87; Conservative 0; Mismatches 91; Indels 0; Gaps 0;
;
; QY 2 ATTCGGACAGTATGAATACAGCGGATTAATACAGGTAAAGTTTCATTAACCGGAAAACC 61
; Db 761222 AATTCGTCAAGATGTATTAAACGATTTTATACCGTCATTTTTTAAATGCTGTTATTCG 761281
;
; QY 62 TTTTAAAGATAATATGAAAGGTATTACCTTTGTCACCAAGGGAATGGTAAATATGCCCG 121
; Db 761282 AGAAAAAATTAATATGAGTCGTCGTCACCTTTGCTATTGAACTTTTGAAGAGGTAA 761341
;
; QY 122 AGTTTTTCACTGAATAGCGAATCCAGCCATTTCTATTTCATATTTCGATGGATGGCTGA 179
; Db 761342 AGACTTAGTTTTTCACTGCAACTTTCGAGGTTTATCCAGAGTAAATATACAGGTTTA 761399
;
;
; RESULT 4
; US-09-643-990A-1
; Sequence 1, Application US/09643990A
; Patent No. 6528289
; GENERAL INFORMATION:
; APPLICANT: Robert D. Fleischmann
; Mark D. Adams
; Owen White
; Hamilton O. Smith
; J. Craig Venter
;
; TITLE OF INVENTION: The Nucleotide sequence of
; the Haemophilus influenzae Rd Genome, Fragments
; Thereof, and Uses Thereof
;
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: MD
; COUNTRY: USA
; ZIP: 20850
;
; COMPUTER READABLE FORM:
```

```
;
;
; NUMBER OF SEQUENCES: 96
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARNOLD, WHITE & DURKEE
; STREET: P.O. BOX 4433
; CITY: HOUSTON
; STATE: TEXAS
; COUNTRY: USA
; ZIP: 77210-4433
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word 6.0 / ASCII text output
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/836.261A
; FILING DATE: 25 Apr 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP95/04264
; FILING DATE: 30 Oct 1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 94870171.9
; FILING DATE: 28 Oct 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: KAMMERER, PATRICIA A.
; REGISTRATION NUMBER: 29,775
; REFERENCE/DOCKET NUMBER: INNS:005
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2047 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 864..2016
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 864..2013
;
; US-08-836-261A-1
;
; Query Match 15.8%; Score 34.2; DB 3; Length 2047;
; Best Local Similarity 49.2%; Pred. No. 0.35;
; Matches 90; Conservative 0; Mismatches 93; Indels 0; Gaps 0;
;
; QY 34 AAGGTAAGTTCATTACACAGGAAACCTTTAAAGAAATAATATGAAAGGTATTACCTTGT 93
; Db 824 AATTTTATATAATTTTAAATAATAATTTTCTAGGAAAAAATGCAAAGCATCATCTTA 883
;
; QY 94 TTGCCAAGGGAATGGTAAATATGCCGAGTTTTCAGTGAATAGCGAATCCAGCCATTT 153
; Db 884 TAGGCAAGCCAAATGTTGGAAATCAAGTCCTTTTATAGAAATGCGAAGCAAGAAATAG 943
;
; QY 154 CTATTTCATATTGACTGGATGCTGAATGTCGACTTTTATAGATAATGACGATGAAGATTT 213
; Db 944 CTATTACAGTCATATTTCAGGTACACTAGATACAGAAATAAAGCCAAATTCATATTC 1003
;
; QY 214 AAT 216
; Db 1004 ATT 1006
;
;
; RESULT 3
; US-09-557-884-1
; Sequence 1, Application US/09557884
; Patent No. 6506581
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: The Nucleotide sequence of
; the Haemophilus influenzae Rd Genome, Fragments
; Thereof, and Uses Thereof
;
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fleischmann et al.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: MD
; COUNTRY: USA
; ZIP: 20850
;
; COMPUTER READABLE FORM:
```

MEDIUM TYPE: 3 1/2 inch diskette  
COMPUTER: Dell Pentium  
OPERATING SYSTEM: MS DOS V6.22  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/643,990A  
FILING DATE: 23-Aug-2000  
CLASSIFICATION: <unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/487,429  
FILING DATE: 1995-06-07  
APPLICATION NUMBER: 08/426,787  
FILING DATE: 1995-04-21  
ATTORNEY/AGENT INFORMATION:  
NAME: Kenley K. Hoover  
REGISTRATION NUMBER: 40,302  
REFERENCE/DOCKET NUMBER: PB186PIC1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 301-610-5790  
TELEFAX: 310-309-8439  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1830121 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-643-990A-1

Query Match 14.9%; Score 32.4; DB 4; Length 1830121;  
Best Local Similarity 48.9%; Pred. No. 8.2;  
Matches 87; Conservative 0; Mismatches 91; Indels 0; Gaps 0;  
  
QY 2 ATTCCGACAGTATGATACAGCGGATTAACAGGTAAGTTTCATTACACGGAACCC 61  
DB 761222 AATTCGTCAAGATGATTAACGATTTATACCAGTCATTTTAAATGCTGTTATGC 761281  
  
QY 62 TTTAAGAATAATATGAAGGTATTACCTTGTTGCCAACGGGAATGTAATATGCCCG 121  
DB 761282 AGAAAAAATTAATATTCAGGTGCGTCAACATTTTGCTATTGAACATTTTGAAGAGGTAA 761341  
  
QY 122 AGTTTTCACATGATAGGATCAACGATTCATTCATATTTGACTGGATGCTGA 179  
DB 761342 AGACTTAGTTTCTAGTCACTGCACTTTTCGAGGTTTATCCAGAGTTAAATACAGGTTTA 761399

RESULT 5  
US-08-916-421B-1/c  
Sequence 1, Application US/08/916,421B  
Patent No. 6503729  
GENERAL INFORMATION:  
APPLICANT: Bult et al.  
TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococcus  
Patent No. 6503729  
TITLE OF INVENTION: jannaschii  
FILE REFERENCE: PB275  
CURRENT APPLICATION NUMBER: US/08/916,421B  
CURRENT FILING DATE: 1997-08-22  
PRIOR APPLICATION NUMBER: US 60/024,428  
PRIOR FILING DATE: 1996-08-22  
NUMBER OF SEQ ID NOS: 3  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 1  
LENGTH: 1664976  
TYPE: DNA  
ORGANISM: Methanococcus jannaschii  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (28222)..(28222)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (28257)..(28258)  
OTHER INFORMATION: n equals a, t, c, or g

NAME/KEY: misc feature  
LOCATION: (84773)..(84773)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (84808)..(84808)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (84812)..(84812)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (98120)..(98120)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (98159)..(98159)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (98239)..(98239)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (98266)..(98266)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (98343)..(98343)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (103998)..(103998)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (148948)..(148948)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (163385)..(163385)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (191989)..(191989)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (191995)..(191995)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (231980)..(231980)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (234187)..(234187)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (234220)..(234220)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (234814)..(234814)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (309398)..(309398)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (309418)..(309418)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (312837)..(312837)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (312993)..(312993)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (319226)..(319226)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (359167)..(359167)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (559241)..(559241)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature

```

; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (779676)..(779676)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (855539)..(855539)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (871619)..(871619)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1084830)..(1084830)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1096846)..(1096846)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1119881)..(1119881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1130881)..(1130881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1310988)..(1310988)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1313224)..(1313224)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1349473)..(1349473)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1349491)..(1349491)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1470091)..(1470091)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1569020)..(1569020)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1602912)..(1602912)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1603734)..(1603734)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1637998)..(1637998)

; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1664854)..(1664855)
; OTHER INFORMATION: n equals a, t, c, or g
; US-08-916-421B-1
Query Match 14.7%; Score 32; DB 4; Length 1664976;
Best Local Similarity 65.3%; Pred. No. 11;
Matches 47; Conservative 0; Mismatches 25; Indels 0; Gaps 0;
QY 24 GGATTAATACAGGTAAAGTTCATTACACGGAAACCTTTAAAGATAATATGAAGCT 83
Db 9533339 GGATTTACATTAGGTGATTCATTAAATAAAGAAATTAAGAGTAAATATGGAAGGA 953280
QY 84 ATTACCTTCTTT 95
Db 953279 GAAAGCTTATTT 953268

RESULT 6
US-08-090-523-27/c
; Sequence 27, Application US/08090523
; Patent No. 5498830
; GENERAL INFORMATION:
; APPLICANT: Barry, Gerard F.
; APPLICANT: Kishore, Ganesh M.
; APPLICANT: Stark, David M.
; TITLE OF INVENTION: Enhanced Starch Biosynthesis
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Grace L. Bonner, Monsanto Co. B34F
; STREET: 700 Chesterfield Parkway No. 5498830th
; CITY: St. Louis
; STATE: Missouri
; COUNTRY: USA
; ZIP: 63198
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/090,523
; FILING DATE: 19930712
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/709663
; FILING DATE: 07-JUN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/539763
; FILING DATE: 18-JUN-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Bonner, Grace L.
; REGISTRATION NUMBER: 32,963
; REFERENCE/DOCKET NUMBER: 38-21(10559)A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (314) 537-7286
; TELEFAX: (314) 537-6047
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1478 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-090-523-27
Query Match 14.6%; Score 31.6; DB 1; Length 1478;
Best Local Similarity 49.4%; Pred. No. 1.9;
Matches 82; Conservative 0; Mismatches 84; Indels 0; Gaps 0;
QY 2 ATTCGGACAGTATGATACGCGGATTATACAGGTAAATACAGGTATTCATTACACGAAACCT 61
```



Db 615 AATCTTACATTAATAAATAACATATAAATAATATCCTTTATTTTGTAAATTTTGAATATGT 556  
QY 62 TTTAAAGAAATATGAAAGGTATTACCTTGTGTCACCGGAATGGTAAATATGCCCG 121  
Db 555 TATGTAAGTATTATTAATAATTTATCGATTTTTTAAAAAGTGAATTGAAATTTTAGTG 496  
QY 122 AGTTTTCCTGTAAGCAATCCAGCCATTTCTATTATTTGA 167  
Db 495 AGAGAAACAATGAAGCAAGCAACTTGATATATATACCTGA 450

RESULT 7  
US-08-398-627-27/c  
; Sequence 27, Application US/08398627  
; Patent No. 5608149  
; GENERAL INFORMATION:  
; APPLICANT: Barry, Gerard F.  
; APPLICANT: Kishore, Ganesh M.  
; APPLICANT: Stark, David M.  
; TITLE OF INVENTION: Enhanced Starch Biosynthesis  
; NUMBER OF SEQUENCES: 51  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Grace L. Bonner, Monsanto Co. B4f  
; STREET: 700 Chesterfield Parkway No. 5608149th  
; CITY: St. Louis  
; STATE: Missouri  
; COUNTRY: USA  
; ZIP: 63198

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/398,627  
; FILING DATE: 03-MAR-1995  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/090,523  
; FILING DATE: 12-JUL-1993  
; APPLICATION NUMBER: US 07/709663  
; FILING DATE: 07-JUN-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/539763  
; FILING DATE: 18-JUN-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bonner, Grace L.  
; REGISTRATION NUMBER: 32,963  
; REFERENCE/DOCKET NUMBER: 38-21(10559)A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (314) 537-7286  
; TELEFAX: (314) 537-6047  
; INFORMATION FOR SEQ ID NO: 27:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1478 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-398-627-27

Query Match 14.6%; Score 31.6; DB 1; Length 1478;  
Best Local Similarity 49.4%; Pred. No. 1.9;  
Matches 82; Conservative 0; Mismatches 84; Indels 0; Gaps 0;  
QY 2 ATTCGGACAGTATGAATACAGCGGATTATACAGGTAAAGTTCAATACACGGGAAAAACC 61  
Db 615 AATCTTACATTAATAAATAACATATAAATAATATCCTTTATTTTGTAAATTTTGAATATGT 556  
QY 62 TTTAAAGAAATATGAAAGGTATTACCTTGTGTCACCGGAATGGTAAATATGCCCG 121  
Db 555 TATGTAAGTATTATTAATAATTTATCGATTTTTTAAAAAGTGAATTGAAATTTTAGTG 496

QY 122 AGTTTTCCTGTAAGCAATCCAGCCATTTCTATTATTTGA 167  
Db 495 AGAGAAACAATGAAGCAAGCAACTTGATATATATACCTGA 450

RESULT 8  
US-08-406-857-1/c  
; Sequence 1, Application US/08406857  
; Patent No. 5608150  
; GENERAL INFORMATION:  
; APPLICANT: Conner, Timothy W.  
; TITLE OF INVENTION: Fruit Specific Promoters  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Grace L. Bonner, Monsanto Company, B4f  
; STREET: 700 Chesterfield Parkway No. 5608150th  
; CITY: St. Louis  
; STATE: Missouri  
; COUNTRY: USA  
; ZIP: 63198  
COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/406,857  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US94/07072  
; FILING DATE: 27-JUN-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/090,523  
; FILING DATE: 12-JUL-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bonner, Grace L.  
; REGISTRATION NUMBER: 32,963  
; REFERENCE/DOCKET NUMBER: 38-21(10655)A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (314) 537-7286  
; TELEFAX: (314) 537-6047  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1478 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-406-857-1

Query Match 14.6%; Score 31.6; DB 1; Length 1478;  
Best Local Similarity 49.4%; Pred. No. 1.9;  
Matches 82; Conservative 0; Mismatches 84; Indels 0; Gaps 0;  
QY 2 ATTCGGACAGTATGAATACAGCGGATTATACAGGTAAAGTTCAATACACGGGAAAAACC 61  
Db 615 AATCTTACATTAATAAATAACATATAAATAATATCCTTTATTTTGTAAATTTTGAATATGT 556  
QY 62 TTTAAAGAAATATGAAAGGTATTACCTTGTGTCACCGGAATGGTAAATATGCCCG 121  
Db 555 TATGTAAGTATTATTAATAATTTATCGATTTTTTAAAAAGTGAATTGAAATTTTAGTG 496  
QY 122 AGTTTTCCTGTAAGCAATCCAGCCATTTCTATTATTTGA 167  
Db 495 AGAGAAACAATGAAGCAAGCAACTTGATATATATACCTGA 450

RESULT 9  
US-08-596-024-3/c  
; Sequence 3, Application US/08596024  
; Patent No. 5716837  
; GENERAL INFORMATION:

APPLICANT: Barry, Gerard F.  
APPLICANT: Deweerdt, Jan W.  
APPLICANT: Kishore, Ganesh M.  
APPLICANT: Weldon, Marcia L.  
TITLE OF INVENTION: Expression of Sucrose Phosphorylase in  
TITLE OF INVENTION: Plants  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Charles E. Cohen, Monsanto Company, BB4F  
STREET: 700 Chesterfield Parkway No. 5716837th  
CITY: St. Louis  
STATE: Missouri  
COUNTRY: USA  
ZIP: 63198

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/596,024  
FILING DATE: US/08/596,024

CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/386,860  
FILING DATE: 10-FEB-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Cohen, Charles E.  
REGISTRATION NUMBER: 34,565  
REFERENCE/DOCKET NUMBER: 38-21(113567)A

TELEPHONE: (314) 537-6224  
TELEFAX: (314) 537-6047  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1478 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-596-024-3

Query Match 14.6%; Score 31.6; DB 1; Length 1478;  
Best Local Similarity 49.4%; Pred. No. 1.9;  
Matches 82; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

QY 2 ATTCGACAGTATGAATACACGGGATTAATACAGGTAAGTTTCATTACACGGAAACC 61  
Db 615 AATCTTACATTAAATAAATATATCCATTTTAAAGAGTAATGAAATTTAGT 556  
QY 62 TTAAAGAAATATATGAAGGTAATACCTTGTTCACCAACGGGAAATGTAATATGCCCG 121  
Db 555 TATGTAAGTATATATAAATATATCCATTTTAAAGAGTAATGAAATTTAGT 496  
QY 122 AGTTTTCAGTAATGAGGTAATCCAGCATTTCTTATTCATATTTGA 167  
Db 495 AGAGAAACAATGACGAAGACGACAACTTGATATATATACCTGA 450

RESULT 10  
US-09-020-818-3/c  
Sequence 3, Application US/09020818  
Patent No. 6222098  
GENERAL INFORMATION:  
APPLICANT: Barry, Gerard F.  
APPLICANT: Deweerdt, Jan W.  
APPLICANT: Kishore, Ganesh M.  
APPLICANT: Weldon, Marcia L.  
TITLE OF INVENTION: Expression of Sucrose Phosphorylase in Plants  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Grace L. Bonner, Monsanto Company, BB4F  
STREET: 700 Chesterfield Parkway No. 6222098th

CITY: St. Louis  
STATE: Missouri  
COUNTRY: USA  
ZIP: 63198  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/020,818  
FILING DATE: February 9, 1998  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Bonner, Grace L.  
REGISTRATION NUMBER: 32,963  
REFERENCE/DOCKET NUMBER: 38-21(10557)A

TELEPHONE: (314) 694-3165  
TELEFAX: (314) 694-9009  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1478 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-09-020-818-3

Query Match 14.6%; Score 31.6; DB 3; Length 1478;  
Best Local Similarity 49.4%; Pred. No. 1.9;  
Matches 82; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

QY 2 ATTCGACAGTATGAATACACGGGATTAATACAGGTAAGTTTCATTACACGGAAACC 61  
Db 615 AATCTTACATTAAATAAATATATCCATTTTAAAGAGTAATGAAATTTAGT 556  
QY 62 TTAAAGAAATATATGAAGGTAATACCTTGTTCACCAACGGGAAATGTAATATGCCCG 121  
Db 555 TATGTAAGTATATATAAATATATCCATTTTAAAGAGTAATGAAATTTAGT 496  
QY 122 AGTTTTCAGTAATGAGGTAATCCAGCATTTCTTATTCATATTTGA 167  
Db 495 AGAGAAACAATGACGAAGACGACAACTTGATATATATACCTGA 450

RESULT 11  
US-08-907-740-3/c  
Sequence 3, Application US/08907740  
Patent No. 6235971  
GENERAL INFORMATION:  
APPLICANT: Barry, Gerard F.  
APPLICANT: Deweerdt, Jan W.  
APPLICANT: Kishore, Ganesh M.  
APPLICANT: Weldon, Marcia L.  
TITLE OF INVENTION: Expression of Sucrose Phosphorylase in  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ARNOLD, WHITE & DURKEE  
STREET: P.O. Box 4433  
CITY: Houston  
STATE: TX  
COUNTRY: U.S.A.  
ZIP: 77210-4433  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/907,740  
FILING DATE:

CLASSIFICATION: 800  
PRIOR APPLICATION DATA: US 08/596,024  
FILING DATE: 06-FEB-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Patterson, Melinda L.  
REGISTRATION NUMBER: 33,062  
REFERENCE/DOCKET NUMBER: MOST171/PAT/HUN  
TELEPHONE: 713-787-1400  
TELEFAX: 713-789-2679  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1478 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-907-467-3

Query Match 14.6%; Score 31.6; DB 3; Length 1478;  
Best Local Similarity 49.4%; Pred. No. 1.9;  
Matches 82; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

QY 2 ATTCGGACAGTATGAATACAGCGGATTATACAGGTAGTTCATTACACGGAAAAACC 61  
Db 615 AATCTTACATTAATAAATAACATATAAAAAATATCCTTTATTTTGAATATGT 556

QY 62 TTAAAGAAATATATGAAAGGTATTACCTTGTTCGCCAAGCGGAATGTAATATGCCCG 121  
Db 555 TAATGAAGTTATATATAAATTTATCGTTTTTTAAAGTAAGTAATTTAGT 496

QY 122 AGTTTTCAGTAAAGGAAATCCAGCCATTTCTATTCATATTGA 167  
Db 495 AGAGAAACAATGAACGAGACGACAACTTGATATATATACCTGA 450

RESULT 12  
US-09-797-467-3/c  
Sequence 3, Application US/09797467  
Patent No. 6476295  
GENERAL INFORMATION:  
APPLICANT: Barry, Gerard F.  
APPLICANT: DeWeerd, Jan W.  
APPLICANT: Kishore, Ganesh M.  
APPLICANT: Welton, Marcia L.  
TITLE OF INVENTION: Expression of Sucrose Phosphorylase in Plants  
FILE REFERENCE: 11899.0004.CNUS02 MOST:004--2  
CURRENT APPLICATION NUMBER: US/09/797,467  
CURRENT FILING DATE: 2001-03-01  
PRIOR APPLICATION NUMBER: US 09/020,818  
PRIOR FILING DATE: 1998-02-09  
PRIOR APPLICATION NUMBER: US 08/596,024  
PRIOR FILING DATE: 1997-02-06  
PRIOR APPLICATION NUMBER: US 08/386,860  
PRIOR FILING DATE: 1995-02-10  
NUMBER OF SEQ ID NOS: 5  
SOFTWARE: Patentin version 3.0  
SEQ ID NO 3  
LENGTH: 1478  
TYPE: DNA  
ORGANISM: Genomic  
US-09-797-467-3

Query Match 14.6%; Score 31.6; DB 4; Length 1478;  
Best Local Similarity 49.4%; Pred. No. 1.9;  
Matches 82; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

QY 2 ATTCGGACAGTATGAATACAGCGGATTATACAGGTAGTTCATTACACGGAAAAACC 61  
Db 615 AATCTTACATTAATAAATAACATATAAAAAATATCCTTTATTTTGAATATGT 556

QY 62 TTAAAGAAATATATGAAAGGTATTACCTTGTTCGCCAAGCGGAATGTAATATGCCCG 121

Db 555 TATGTAAGTTATATATAAATTTATCGATTTTAAAAAGTGAATTTAGT 496

QY 122 AGTTTTCAGTAAAGGAAATCCAGCCATTTCTATTCATATTGA 167

Db 495 AGAGAAACAATGAACGAGACGACAACTTGATATATATACCTGA 450

RESULT 13  
US-08-399-023-27/c  
Sequence 27, Application US/08399023  
Patent No. 6538179  
GENERAL INFORMATION:  
APPLICANT: Barry, Gerard F.  
APPLICANT: Kishore, Ganesh M.  
APPLICANT: Stark, David M.  
TITLE OF INVENTION: Enhanced Starch Biosynthesis  
NUMBER OF SEQUENCES: 51  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Grace L. Bonner, Monsanto Co. BB4F  
STREET: 700 Chesterfield Parkway No. 6538179th  
CITY: St. Louis  
STATE: Missouri  
COUNTRY: USA  
ZIP: 63198  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/399,023  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/090,523  
FILING DATE:  
APPLICATION NUMBER: US 07/709663  
FILING DATE: 07-JUN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/539763  
FILING DATE: 18-JUN-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Bonner, Grace L.  
REGISTRATION NUMBER: 32,963  
REFERENCE/DOCKET NUMBER: 38-21(10559)A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (314) 537-7286  
TELEFAX: (314) 537-6047  
INFORMATION FOR SEQ ID NO: 27:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1478 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-399-023-27

Query Match 14.6%; Score 31.6; DB 4; Length 1478;  
Best Local Similarity 49.4%; Pred. No. 1.9;  
Matches 82; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

QY 2 ATTCGGACAGTATGAATACAGCGGATTATACAGGTAGTTCATTACACGGAAAAACC 61  
Db 615 AATCTTACATTAATAAATAACATATAAAAAATATCCTTTATTTTGAATATGT 556

QY 62 TTAAAGAAATATATGAAAGGTATTACCTTGTTCGCCAAGCGGAATGTAATATGCCCG 121

Db 555 TAATGAAGTTATATATAAATTTATCGATTTTAAAAAGTGAATTTAGT 496

QY 122 AGTTTTCAGTAAAGGAAATCCAGCCATTTCTATTCATATTGA 167

Db 495 AGAGAAACAATGAACGAGACGACAACTTGATATATATACCTGA 450

QY 72 AATATGAAAGGTATTACCTTGTTT 95  
Db 2385 TTTATATATAAATCTTACATTTT 2362

Search completed: November 15, 2003, 08:09:51  
Job time : 42.7995 secs

RESULT 14  
PCT-US94-07072-1/c  
; Sequence 1, Application PC/TUS9407072  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: Fruit Specific Promoters  
; NUMBER OF SEQUENCES: 2  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US94/07072  
; FILING DATE: 27-JUN-1994  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/090523  
; FILING DATE: 12-JUL-1993  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1478 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
PCT-US94-07072-1

Query Match 14.6%; Score 31.6; DB 5; Length 1478;  
Best Local Similarity 49.4%; Pred. NO. 1.9;  
Matches 82; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

QY 2 ATTCGACAGTATGATACACGGGATTATACAGGTAAGTTCATTACACGGGAAACC 61  
Db 615 AATCTTACATTAATAAATAACATATAAAATATCCTTTATTTGTAATTTGATATGT 556

QY 62 TTTAAAGAAATATATGAAGGTATTACCTGTTGCCAACGGGAATGTAATAATGCCCG 121  
Db 555 TATGTAAGATTATATTAATAATTTATCGATTTTTTAAAAAGTGAATGAAAAATTTAGTG 496

QY 122 AGTTTTTCATCAATAGCAATCCAGCCATTCATTCATATTTGA 167  
Db 495 AGAGAAACAATGACGACGAAGACGACAACTTGATATATATACCTGA 450

RESULT 15  
US-09-328-352-4115/c  
; Sequence 4115, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; FILE REFERENCE: GTC99-03PA  
; CURRENT APPLICATION NUMBER: US/09/328,352  
; CURRENT FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 4115  
; LENGTH: 3030  
; TYPE: DNA  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-4115

Query Match 14.4%; Score 31.2; DB 4; Length 3030;  
Best Local Similarity 60.7%; Pred. NO. 3.1;  
Matches 51; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 12 TATGATACAGCGGATTATACAGGTAAGTTCATTACACGGGAAACCTTTAAAGAAT 71  
Db 2445 TATGAACATCATCTATGTATAAATAAATTTTATAGACTTACAACTCTTTAAACT 2386

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 15, 2003, 00:09:34 ; Search time 122.893 Seconds  
(without alignments)  
5771.357 Million cell updates/sec

Title: US-09-928-457-99

Perfect score: 217

Sequence: 1 AATTCGACAGTATGATAC.....ATGACGATGAAGATTAAATT 217

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2169961 seqs, 1634102185 residues

Total number of hits satisfying chosen parameters: 4339922

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:\*
- 12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*
- 13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:\*
- 14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:\*
- 15: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*
- 16: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*
- 17: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	217	100.0	217	10	US-09-928-457-99
2	37.6	17.3	507	11	Sequence 99, Appl
3	37.6	17.3	602	10	Sequence 22577, A
4	37.6	17.3	2998	9	Sequence 730, App
5	37.6	17.3	3920	14	Sequence 295, App
6	37.6	17.3	5144	11	Sequence 133, App
7	36	16.6	2000	10	Sequence 70, Appl
8	36	16.6	2004	10	Sequence 3712, Ap
9	36	16.6	2003	10	Sequence 306, App
10	36	16.6	640681	10	Sequence 253, App
11	34.8	16.0	17421	12	Sequence 1, Appl
12	34.8	16.0	17421	14	Sequence 56, Appl
13	33	15.2	334	10	Sequence 54, Appl
14	33	15.2	495	11	Sequence 19, Appl
15	32.8	15.1	18434	12	Sequence 22573, A
16	32.6	15.0	15881	12	Sequence 1979, Ap
					Sequence 234, App

c 17	32.4	14.9	585	12	US-10-027-632-220246	Sequence 220246,
c 18	32.4	14.9	585	13	US-10-027-632-220246	Sequence 220246,
c 19	32.4	14.9	1830121	14	US-10-329-960-1	Sequence 1, Appli
c 20	32.2	14.8	2645	12	US-10-027-632-252267	Sequence 252267,
c 21	32.2	14.8	2645	13	US-10-027-632-252267	Sequence 252267,
c 22	32	14.7	7143	12	US-10-311-455-956	Sequence 956, App
c 23	31.8	14.7	6436	12	US-10-311-455-1665	Sequence 1665, Ap
c 24	31.6	14.6	427	10	US-09-878-178-449	Sequence 449, App
c 25	31.6	14.6	427	13	US-10-046-935-449	Sequence 449, App
c 26	31.6	14.6	427	14	US-10-146-503-449	Sequence 449, App
c 27	31.6	14.6	516	12	US-10-027-632-61785	Sequence 61785, A
c 28	31.6	14.6	516	12	US-10-027-632-61786	Sequence 61786, A
c 29	31.6	14.6	516	13	US-10-027-632-61785	Sequence 61785, A
c 30	31.6	14.6	1478	9	US-10-027-632-61786	Sequence 61786, A
c 31	31.6	14.6	1478	14	US-10-287-933-3	Sequence 3, Appli
c 32	31.6	14.6	2021	9	US-09-822-849A-233	Sequence 233, App
c 33	31.6	14.6	32205	11	US-09-764-891-10213	Sequence 10213, A
c 34	31.6	14.6	32205	14	US-10-205-428-10123	Sequence 1012, Ap
c 35	31.6	14.5	578	12	US-10-027-632-44395	Sequence 44395, A
c 36	31.4	14.5	578	12	US-10-027-632-44396	Sequence 44396, A
c 37	31.4	14.5	578	13	US-10-027-632-44395	Sequence 44395, A
c 38	31.4	14.5	578	13	US-10-027-632-44396	Sequence 44396, A
c 39	31.4	14.5	578	13	US-10-027-632-44396	Sequence 44396, A
c 40	31.4	14.5	2236	12	US-09-814-353-19364	Sequence 19364, A
c 41	31.4	14.5	2241	14	US-10-198-845-13592	Sequence 13592, A
c 42	31.2	14.4	445	10	US-09-923-300-8	Sequence 8, Appli
c 43	31.2	14.4	786431	12	US-10-412-277-3	Sequence 3, Appli
c 44	31	14.3	481	11	US-09-918-995-399	Sequence 399, App
c 45	31	14.3	5763	12	US-10-311-455-156	Sequence 156, App

ALIGNMENTS

RESULT 1

US-09-928-457-99

; Sequence 99, Application US/09928457

; Patent No. US20020164603A1

; GENERAL INFORMATION:

; APPLICANT:

; TITLE OF INVENTION: DNA, specific proteins and peptides

; TITLE OF INVENTION: of the Neisseria meningitidis species bacteria, method

; NUMBER OF SEQUENCES: 99

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30 (OEB)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/928,457

; FILING DATE: 2001-08-14

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 09/214,759

; FILING DATE: 199-12-10

; INFORMATION FOR SEQ ID NO: 99:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 217 base pairs

; TYPE: nucleotide

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; HYPOTHEICAL: NO

; ANTI-SENSE: NO

US-09-928-457-99

Query Match 100.0%; Score 217; DB 10; Length 217;

Best Local Similarity 100.0%; Pred. No. 1.4e-49;

Matches 217; Conservative 0; Mismatches 0; Gaps 0;

QY 1 AATTCGACAGTATGATACAGCGATTAAATACAGGTAAGTTCATTACACGGAAC 60

DB 1 AATTCGACAGTATGATACAGCGATTAAATACAGGTAAGTTCATTACACGGAAC 60

```
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-919-580-730

Query Match      17.3%; Score 37.6; DB 10; Length 602;
Best Local Similarity 51.8%; Pred. No. 1.8;
Matches      85; Conservative      0; Mismatches      79; Indels      0; Gaps      0;

QY      1  AATTCGGACAGTATGAATACACGCGGATTAATACAGGTAAGTTCAATACACGGAATAAC 60
Db      1  AATTCGGACAGTATGAATACACGCGGATTAATACAGGTAAGTTCAATACACGGAATAAC 60
QY      417 AACTGAGGCAATTAATACCTCTTTATATAATGCAAGTTGAAATGCTAACAAAGCATAAAC 358
Db      417 AACTGAGGCAATTAATACCTCTTTATATAATGCAAGTTGAAATGCTAACAAAGCATAAAC 358
QY      61  CTTTAAAGAAATATGAAAGGTATTACCTTGTTTCCCAACGGGAATGGTAAATATGCCC 120
Db      61  CTTTAAAGAAATATGAAAGGTATTACCTTGTTTCCCAACGGGAATGGTAAATATGCCC 120
QY      357 ACTTCTGCAAAAATTCACAGGCGACAGTTGTTCAATCAACAGAAAAGTCAAAACCACT 298
Db      357 ACTTCTGCAAAAATTCACAGGCGACAGTTGTTCAATCAACAGAAAAGTCAAAACCACT 298
QY      121 GAGTTTTCACCTGAATAGCAATCCAGGCATTTCTATTCAATTT 164
Db      121 GAGTTTTCACCTGAATAGCAATCCAGGCATTTCTATTCAATTT 164
QY      297 TGGTTTTTAAATGAAAATCCCTTCACATCCACCTGTGTTCAAAAT 254
Db      297 TGGTTTTTAAATGAAAATCCCTTCACATCCACCTGTGTTCAAAAT 254

RESULT 4
US-09-925-302-295/c
; Sequence 295, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 295
; LENGTH: 2998
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (11)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (16)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (195)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (2967)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (2971)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (2977)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (2981)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-302-295

Query Match      17.3%; Score 37.6; DB 9; Length 2998;
Best Local Similarity 51.8%; Pred. No. 3.3;
Matches      85; Conservative      0; Mismatches      79; Indels      0; Gaps      0;

QY      1  AATTCGGACAGTATGAATACACGCGGATTAATACAGGTAAGTTCAATACACGGAATAAC 60
```

```
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-22577/c
; Sequence 22577, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 22577
; LENGTH: 507
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(507)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-22577

Query Match      17.3%; Score 37.6; DB 11; Length 507;
Best Local Similarity 51.8%; Pred. No. 1.7;
Matches      85; Conservative      0; Mismatches      79; Indels      0; Gaps      0;

QY      1  AATTCGGACAGTATGAATACACGCGGATTAATACAGGTAAGTTCAATACACGGAATAAC 60
Db      1  AATTCGGACAGTATGAATACACGCGGATTAATACAGGTAAGTTCAATACACGGAATAAC 60
QY      497 AACTGAGGCAATTAATACCTCTTTATATAATGCAAGTTGAAATGCTAACAAAGCATAAAC 438
Db      497 AACTGAGGCAATTAATACCTCTTTATATAATGCAAGTTGAAATGCTAACAAAGCATAAAC 438
QY      61  CTTTAAAGAAATATGAAGGTATTACCTTGTTTCCCAACGGGAATGGTAAATATGCCC 120
Db      61  CTTTAAAGAAATATGAAGGTATTACCTTGTTTCCCAACGGGAATGGTAAATATGCCC 120
QY      437 GCTTCTGCAAAAATTCACAGGCGACAGTTGTTCAATCAACAGAAAAGTCAAAACCACT 378
Db      437 GCTTCTGCAAAAATTCACAGGCGACAGTTGTTCAATCAACAGAAAAGTCAAAACCACT 378
QY      121 GAGTTTTCACCTGAATAGCAATCCAGGCATTTCTATTCAATTT 164
Db      121 GAGTTTTCACCTGAATAGCAATCCAGGCATTTCTATTCAATTT 164
QY      377 TGGTTTTTAAATGAAAATCCCTTCACATCCACCTGTGTTCAAAAT 334
Db      377 TGGTTTTTAAATGAAAATCCCTTCACATCCACCTGTGTTCAAAAT 334

RESULT 3
US-09-919-580-730/c
; Sequence 730, Application US/09919580
; Patent No. US2002010832A1
; GENERAL INFORMATION:
; APPLICANT: Pyle, Ruth
; APPLICANT: Xu, Jiangchun
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121,552
; CURRENT APPLICATION NUMBER: US/09/919,580
; CURRENT FILING DATE: 2001-07-30
; NUMBER OF SEQ ID NOS: 934
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 730
; LENGTH: 602
```

Db 2282 AACTGAGCAATTAATACCTCTTTTATAATGAAAGTTGAAATGCTAAACAAGCATAAAC 2223  
Qy 61 CTTTAAAGAATAATATGAAGGTATTACCTCTTTTCCAAACGGGAATGGTAAATATGCC 120  
Db 2222 ACTCTGCAAAAATCCCAAGGCACAGTTGTTTCATTCAACAGAAAAGTCAAAACCACT 2163  
Qy 121 GAGTTTTCACCTGAATAGGAATCCAGGCATTTCTTATTCATATT 164  
Db 2162 TGGTTTTTAATGAAATCCITTCACATCCACCTGTTTCAAAAT 2119

RESULT 5  
US-10-084-817-133/c  
; Sequence 133, Application US/10084817  
; Publication No. US20030119009A1  
; GENERAL INFORMATION:  
; APPLICANT: Susan Stuart  
; APPLICANT: Jed G. Nuchtern  
; APPLICANT: Sharon E. Plon  
; APPLICANT: Jason M. Shohet  
; TITLE OF INVENTION: GENES REGULATED BY MYCN ACTIVATION  
; FILE REFERENCE: PA-0046 US  
; CURRENT APPLICATION NUMBER: US/10/084,817  
; CURRENT FILING DATE: 2002-02-25  
; PRIOR APPLICATION NUMBER: 60/270,784  
; PRIOR FILING DATE: 2001-02-23  
; NUMBER OF SEQ ID NOS: 365  
; SOFTWARE: PERL Program  
; SEQ ID NO 133  
; LENGTH: 3920  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No. US20030119009A1 407096.14

US-10-084-817-133  
Query Match 17.3%; Score 37.6; DB 14; Length 3920;  
Best Local Similarity 51.8%; Pred. No. 3.7; Indels 0; Gaps 0;  
Matches 85; Conservative 0; Mismatches 79; Indels 0; Gaps 0;  
Qy 1 AATTCGACAGTATGAATACAGCGGATTAATACAGGTAAGTTCAATTAACAACGGAAAAAC 60  
Db 3277 AACTGAGCAATTAATACCTCTTTTATAATGCAAGTTGAAATGCTAAACAAGCATAAAC 3218  
Qy 61 CTTTAAAGAATAATGAAGGTATTACCTCTTTTCCAAACGGGAATGGTAAATATGCC 120  
Db 3217 ACTCTGCAAAAATCCCAAGGCACAGTTGTTTCATTCAACAGAAAAGTCAAAACCACT 3158  
Qy 121 GAGTTTTCACCTGAATAGGAATCCAGGCATTTCTTATTCATATT 164  
Db 3157 TGGTTTTTAATGAAATCCITTCACATCCACCTGTTTCAAAAT 3114

RESULT 6  
US-09-919-039-70/c  
; Sequence 70, Application US/09919039  
; Publication No. US20030108871A1  
; GENERAL INFORMATION:  
; APPLICANT: Kaser, Matthew R.  
; TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTURES  
; FILE REFERENCE: PA-0035 US  
; CURRENT APPLICATION NUMBER: US/09/919,039  
; CURRENT FILING DATE: 2002-09-09  
; PRIOR APPLICATION NUMBER: 60/222,113  
; PRIOR FILING DATE: 2000-07-28  
; NUMBER OF SEQ ID NOS: 401  
; SOFTWARE: PERL Program  
; SEQ ID NO 70  
; LENGTH: 5144  
; TYPE: DNA  
; ORGANISM: Homo sapiens

; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No. US20030108871A1 407096.2  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: 4537, 4540, 4590, 5133  
; OTHER INFORMATION: a, t, c, g, or other  
US-09-919-039-70

Query Match 17.3%; Score 37.6; DB 11; Length 5144;  
Best Local Similarity 51.8%; Pred. No. 4.1; Indels 0; Gaps 0;  
Matches 85; Conservative 0; Mismatches 79; Indels 0; Gaps 0;  
Qy 1 AATTCGACAGTATGAATACAGCGGATTAATACAGGTAAGTTCAATTAACAACGGAAAAAC 60  
Db 3277 AACTGAGCAATTAATACCTCTTTTATAATGCAAGTTGAAATGCTAAACAAGCATAAAC 3218  
Qy 61 CTTTAAAGAATAATGAAGGTATTACCTCTTTTCCAAACGGGAATGGTAAATATGCC 120  
Db 3217 ACTCTGCAAAAATCCCAAGGCACAGTTGTTTCATTCAACAGAAAAGTCAAAACCACT 3158  
Qy 121 GAGTTTTCACCTGAATAGGAATCCAGGCATTTCTTATTCATATT 164  
Db 3157 TGGTTTTTAATGAAATCCITTCACATCCACCTGTTTCAAAAT 3114

RESULT 7  
US-09-938-842A-3712  
; Sequence 3712, Application US/09938842A  
; Patent No. US20020160378A1  
; GENERAL INFORMATION:  
; APPLICANT: Harper, Jeff  
; APPLICANT: Kreps, Joel  
; APPLICANT: Wang, Xun  
; APPLICANT: Zhu, Tong  
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING  
; FILE REFERENCE: S01P1300-3  
; CURRENT APPLICATION NUMBER: US/09/938,842A  
; CURRENT FILING DATE: 2001-08-24  
; PRIOR APPLICATION NUMBER: US 60/227,866  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: US 60/254,647  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/300,111  
; PRIOR FILING DATE: 2001-06-22  
; NUMBER OF SEQ ID NOS: 5379  
; SEQ ID NO 3712  
; LENGTH: 2000  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
US-09-938-842A-3712

Query Match 16.8%; Score 36; DB 10; Length 2000;  
Best Local Similarity 51.9%; Pred. No. 7.8; Indels 0; Gaps 0;  
Matches 81; Conservative 0; Mismatches 75; Indels 0; Gaps 0;  
Qy 29 AATCAAGGTGAAGTTCAATTAACAACGGAAAAACCTTTAAAGAAATAATGAAGGTATTAC 88  
Db 13 AAAACATCGAACGAAAATTAATACGTAGCACCATAAAATGAGNAAAATACACAACTAT 72  
Qy 89 CTGTGTGGCAACGGGAATGGTAAATATGCCGAGTTTTCATCTGAATAGCGAATCCAGC 148  
Db 73 AGTGATTGACGATGCTTTTGACATATAAGTCTTACTCGTTACAGATTATTAAATCTATT 132  
Qy 149 CATTCTATTCATATTGACTGGATGGCTGAATGTC 184  
Db 133 TTTTTCGATTGATCTTACGTTATGCAAAATATG 168

RESULT 8  
US-09-887-576-306  
; Sequence 306, Application US/09887576

```
; Patent No. US20020144047A1
; GENERAL INFORMATION:
; APPLICANT: Budworth, P.
; APPLICANT: Brown, D.
; APPLICANT: Chang, H.
; APPLICANT: Zhu, T.
; APPLICANT: Han, B.
; APPLICANT: Wang, X.
; APPLICANT: Cooper, Bret
; TITLE OF INVENTION: Promoters for regulation of plant expression
; FILE REFERENCE: 1360.001US1
; CURRENT APPLICATION NUMBER: US/09/887,576
; CURRENT FILING DATE: 2001-06-25
; PRIOR APPLICATION NUMBER: US 60/213,848
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 60/214,087
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 60/258,692
; PRIOR FILING DATE: 2000-12-29
; NUMBER OF SEQ ID NOS: 875
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 306
; LENGTH: 2003
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-887-576-306

Query Match      16.6%; Score 36; DB 10; Length 2003;
Best Local Similarity 51.9%; Pred. No. 7.8;
Matches 81; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 29 AATCAAGTAAAGTTCATTACACGGAACCTTTAAAGAAATATATGAAAGGTATTAC 88
DB 13 AAACATCGACGAAATTAACGTAAAGCACCATAAAATGAGAAAAATACAAACTAT 72
QY 89 CTTGTTGCCAACGGGAATGTAATATGCCCGAGTTTTTCACCTGAATAGCGAATCCAGC 148
DB 73 AGTGATTGACGATGCTTTTGACATATAAGTCTTACTCGTTAACAGATTATAATCTATT 132
QY 149 CATTTCTATTATTTGACCTGGATGCTGAATGTG 184
DB 133 TTTTTCGATTGATCTTACGCTTTATGCAAAATTATG 168

RESULT 9
US-09-887-576-253
; Sequence 253, Application US/09887576
; Patent No. US20020144047A1
; GENERAL INFORMATION:
; APPLICANT: Budworth, P.
; APPLICANT: Brown, D.
; APPLICANT: Chang, H.
; APPLICANT: Zhu, T.
; APPLICANT: Han, B.
; APPLICANT: Wang, X.
; APPLICANT: Cooper, Bret
; TITLE OF INVENTION: Promoters for regulation of plant expression
; FILE REFERENCE: 1360.001US1
; CURRENT APPLICATION NUMBER: US/09/887,576
; CURRENT FILING DATE: 2001-06-25
; PRIOR APPLICATION NUMBER: US 60/213,848
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 60/214,087
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 60/258,692
; PRIOR FILING DATE: 2000-12-29
; NUMBER OF SEQ ID NOS: 875
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 253
; LENGTH: 2004
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-887-576-253
```

```
Query Match      16.6%; Score 36; DB 10; Length 2004;
Best Local Similarity 51.9%; Pred. No. 7.8;
Matches 81; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 29 AATCAAGTAAAGTTCATTACACGGAACCTTTAAAGAAATATATGAAAGGTATTAC 88
DB 13 AAACATCGACGAAATTAACGTAAAGCACCATAAAATGAGAAAAATACAAACTAT 72
QY 89 CTTGTTGCCAACGGGAATGTAATATGCCCGAGTTTTTCACCTGAATAGCGAATCCAGC 148
DB 73 AGTGATTGACGATGCTTTTGACATATAAGTCTTACTCGTTAACAGATTATAATCTATT 132
QY 149 CATTTCTATTATTTGACCTGGATGCTGAATGTG 184
DB 133 TTTTTCGATTGATCTTACGCTTTATGCAAAATTATG 168

RESULT 10
US-09-790-988-1
; Sequence 1, Application US/09790988
; Patent No. US20020127687A1
; GENERAL INFORMATION:
; APPLICANT: SHIGENOBU, SHUJI
; APPLICANT: WATANABE, HIDEMI
; APPLICANT: HATTORI, MASAHIRA
; APPLICANT: SAKAKI, YOSHIYUKI
; TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
; FILE REFERENCE: 081356/0159
; CURRENT APPLICATION NUMBER: US/09/790,988
; CURRENT FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: JP2000-107160
; PRIOR FILING DATE: 2000-04-07
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 640681
; TYPE: DNA
; ORGANISM: Buchnera sp.
US-09-790-988-1

Query Match      16.6%; Score 36; DB 10; Length 640681;
Best Local Similarity 55.6%; Pred. No. 67;
Matches 69; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 27 TTAATACAAAGGTAAGTTCATTACAAACGGAACCTTTAAAGAAATATATGAAAGGTATT 86
DB 512832 TTAATTAAGCATCCAAAGAAATAGGAAATCTAAATGAATAGTATTATTATTATT 512891
QY 87 ACCTTGTTGCCAACGGGAATGTAATATGCCCGAGTTTTTCACCTGAATAGCGAATCCA 146
DB 512892 AAACCTCTTCTTAAAGGACCTCTTAAACAAAGCCGCAAAATTTGCTGGAATACCTAAACCA 512951
QY 147 GCCA 150
DB 512952 GTCA 512955

RESULT 11
US-10-240-453-56/c
; Sequence 56, Application US/10240453
; Publication No. US20030148326A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA
; TITLE OF INVENTION: Transcription
; TITLE OF INVENTION: By Means of Assessing the Methylation Status of Genes Associated
; TITLE OF INVENTION: With DNA Transcription
; FILE REFERENCE: 5013.1009
; CURRENT APPLICATION NUMBER: US/10/240,453
; CURRENT FILING DATE: 2002-10-02
```



```
; PRIOR APPLICATION NUMBER: PCT/EP01/03973
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 350
; SEQ ID NO 56
; LENGTH: 17421
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (167, 1832..1833, 1836..1837, 1842, 1848, 1852, 1857,
; LOCATION: (1859..1860)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (2953, 4024, 4630, 4733, 4767, 8161, 8189, 8200, 9142..9143)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (9147, 9153, 9168, 9170, 9173, 10243, 10340, 10394, 11996)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (12002, 12005, 12012, 12102, 13715, 13717, 16077, 16097)
; US-10-240-453-56
Query Match 16.0%; Score 34.8; DB 12; Length 17421;
Best Local Similarity 65.4%; Pred. No. 37;
Matches 51; Conservative 0; Mismatches 27; Indels 0; Gaps 0;
QY 16 AATACAGCGGATTATACAGGTAGTTCATTACACCGGAAACCTTTAAAGAAATAATA 75
Db 6042 AATATAAAATATTATACCATATAATTTTATTAACCAATAAAACATAAAAAATAATA 5983
QY 76 TGAAGGTATTACCTTGT 93
Db 5982 TATAAATTATATCTTTT 5965
RESULT 12
US-10-239-676-54/c
; Sequence 54, Application US/10239676
; Publication No. US20030082609A1
; GENERAL INFORMATION:
; APPLICANT: GLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
; FILE REFERENCE: 5013.1003
; CURRENT APPLICATION NUMBER: US/10/239,676
; CURRENT FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: PCT/EP01/03968
; DE 10019058.8
; DE 10019173.8
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-04-06
; 2000-04-06
; 2000-04-07
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 54
; LENGTH: 17421
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
```

```
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (167, 1832..1833, 1836..1837, 1842, 1848, 1852, 1857, 1859..1860)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (2953, 4024, 4630, 4733, 4767, 8161, 8189, 8200, 9142..9143)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (9147, 9153, 9168, 9170, 9173, 10243, 10340, 10394, 11996)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (12002, 12005, 12012, 12102, 13715, 13717, 16077, 16097)
; US-10-239-676-54
Query Match 16.0%; Score 34.8; DB 14; Length 17421;
Best Local Similarity 65.4%; Pred. No. 37;
Matches 51; Conservative 0; Mismatches 27; Indels 0; Gaps 0;
QY 16 AATACAGCGGATTATACAGGTAGTTCATTACACCGGAAACCTTTAAAGAAATAATA 75
Db 6042 AATATAAAATATTATACCATATAATTTTATTAACCAATAAAACATAAAAAATAATA 5983
QY 76 TGAAGGTATTACCTTGT 93
Db 5982 TATAAATTATATCTTTT 5965
RESULT 13
US-09-928-457-19/c
; Sequence 19, Application US/09928457
; Patent No. US20020164603A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: DNA, specific proteins and peptides
; TITLE OF INVENTION: of the Neisseria meningitidis species bacteria, method
; TITLE OF INVENTION: for obtaining them and their biological application.
; NUMBER OF SEQUENCES: 99
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (OEB)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/928,457
; FILING DATE: 2001-08-14
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/214,759
; FILING DATE: 199-12-10
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 334 base pairs
; TYPE: nucleotide
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Neisseria meningitidis
; STRAIN: Z2491
; US-09-928-457-19
Query Match 15.2%; Score 33; DB 10; Length 334;
Best Local Similarity 51.0%; Pred. No. 26;
Matches 78; Conservative 0; Mismatches 75; Indels 0; Gaps 0;
QY 35 AGTAAAGTTCATTACACCGGAAACCTTTAAAGAAATAATAATGAAGGTATTACCTTGT 94
Db 327 AGGTGCTATTATGTTTCGAATAATGCTATAACTATAATAGAAAACAATATTACACTAGG 268
QY 95 TGCACACGGGAATGGTAAATATGCCGAGTTTTTCACTGAATAGCGAATCCAGCCATTTC 154
Db 267 TCGGTATAGGATGCTATTATTTCATATTTTCTCATGAAATGTTGTTAGGCGATTGG 208
```

US-10-311-455-1979  
Query Match 15.1%; Score 32.8; DB 12; Length 18434;  
Best Local Similarity 50.6%; Pred. No. 1.3e+02;  
Matches 79; Conservative 0; Mismatches 77; Indels 0; Gaps 0;  
QY 1 AATTCGGACAGTATGAATACACGCGGATTAAATACAAAGGTAAGTTTCATTACAAACGGAAAAAC 60  
DB 2860 AAATACGCCACTACACTCCAACTAAATACAAAAAACCCCTATCTCAAAAAAATAA 2801  
QY 61 CTTTAAAGATAAATATGAAGGTATTACCTTGTTCACACGGGAATGTTAAATATGCCC 120  
DB 2800 AAAAAAATTAATAAATAATATATACCCATAAATAAATAAATAAATAAATAAATAA 2741  
QY 121 GAGTTTTTCACTGAATAGCGGATCCAGCCATTCTTA 156  
DB 2740 TAAATATTTTCAACAACTTACAAATCAACCAATCTCAA 2705  
Search completed: November 15, 2003, 08:32:49  
Job time : 125.983 secs

QY 155 TATTCATATTTGACTGGATGCTGAATGTGGAC 187  
DB 207 TACGATTAACTCAATGTACGAAGATATTGGAC 175  
RESULT 14  
US-09-918-995-22573/c  
; Sequence 22573, Application US/09918995  
; Publication No. US20030073623A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; FILE REFERENCE: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918,995  
; PRIOR FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235,076  
; PRIOR FILING DATE: 1999-01-20  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 22573  
; LENGTH: 495  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(495)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-918-995-22573

Query Match 15.2%; Score 33; DB 11; Length 495;  
Best Local Similarity 49.7%; Pred. No. 30;  
Matches 84; Conservative 0; Mismatches 85; Indels 0; Gaps 0;  
QY 1 AATTCGGACAGTATGAATACACGCGGATTAAATACAAAGGTAAGTTTCATTACAAACGGAAAAAC 60  
DB 478 AACTGAGGCATTAAATACCTCTTTTATAATATGCAAGTTGAATGCTAACAAAGCATAAAC 419  
QY 61 CTTTAAAGATAAATGAAGGTATTACCTTGTTCACACGGGAATGTTAAATATGCCC 120  
DB 418 ACTTCTGCAAAAATTCACAAAGGCACAGTTGTTTCATTCAACAGAAAAAGTCAAAAACCACT 359  
QY 121 GAGTTTTTCACTGAATAGCGGATCCAGCCATTTCATTATATTGACT 169  
DB 358 TGGTTTTTAAATGAATAATCCTTCACCTTGTTCAAAATATTAT 310

QY 155 TATTCATATTTGACTGGATGCTGAATGTGGAC 187  
DB 207 TACGATTAACTCAATGTACGAAGATATTGGAC 175  
RESULT 15  
US-10-311-455-1979/c  
; Sequence 1979, Application US/10311455  
; Publication No. US20030143606A1  
; GENERAL INFORMATION:  
; APPLICANT: OLEK, Alexander  
; APPLICANT: PIEPENBROCK, Christian  
; APPLICANT: BERLIN, Kurt  
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining Cytosine Methylation  
; FILE REFERENCE: 5013.1014  
; CURRENT APPLICATION NUMBER: US/10/311,455  
; CURRENT FILING DATE: 2002-12-16  
; PRIOR APPLICATION NUMBER: PCT/EP01/07537  
; PRIOR FILING DATE: 2001-07-02  
; PRIOR APPLICATION NUMBER: DE 10032529.7  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: DE 10043826.1  
; PRIOR FILING DATE: 2000-09-01  
; NUMBER OF SEQ ID NOS: 2424  
; SEQ ID NO 1979  
; LENGTH: 18434  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)